QUALITY CONTROL CONFERENCE PAPERS 1953

SEVENTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

QUALITY CONTROL CONFERENCE PAPERS 1953

SEVENTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

MAY 27, 28, 29, 1953

PHILADELPHIA, PA.

PRICE

$3.00 At The Conference

$3.50 If Ordered By Mail

Copyright 1953, by
AMERICAN SOCIETY FOR QUALITY CONTROL, INC.
Room 5036
70 East 45th Street
New York 17, New York

Printed in the United States of America
by
THE ENGLE PRESS
Philadelphia, Pa.

FOREWORD

The excellent cooperation of the authors in meeting an early deadline and in presenting their manuscripts in suitable form for direct reproduction has made this publication possible.

While these proceedings are copyrighted, the American Society for Quality Control assumes no responsibility for any of the authors' statements. Responsibility for the context of each paper resides with its author.

Edward M. Schrock
Chairman, Proceedings Committee

NATIONAL OFFICERS

SIMON COLLIER, President
Johns-Manville Corporation

ARTHUR BENDER, JR., Vice President
Delco-Remy Division, GMC

RAYMOND S. SADDORIS, Vice President
A. O. Smith Corporation

JULIAN H. TOULOUSE, Vice President
Owens-Illinois Glass Comapny

EDWARD B. HADEN, Executive Secretary
The Esterbrook Pen Company

PAUL A. ROBERT, Treasurer
International Business Machines Corp.

WADE R. WEAVER, Past President, and Chairman,
General Convention Committee,
Republic Steel Corporation

MASON E. WESCOTT, Chairman, Editorial Board
Rutgers University

PHILADELPHIA SECTION OFFICERS

ALLEN R. CRAWFORD, National Representative
Rutgers University

JOSEPH E. BEEZER, Chairman
RCA Moorestown Engineering Plant

E. G. W. JOHNSTON, Vice Chairman
Ford Motor Company

PAUL A. GEARY, Secretary
Smith, Kline & French Labs.

ANDREW C. IRVINE, Treasurer
The Electric Storage Battery Company

PHILIP I. McGRATH, Editor
U. S. Pipe & Foundry Company

TABLE OF CONTENTS AND CONVENTION PROGRAM

PERIOD I - WEDNESDAY, MAY 27, 2:30 - 4:30 P.M.

A. GENERAL INTEREST

Moderator - Joseph M. Juran, Tuckahoe, N. Y.

WHAT IS THE QUALITY CONTROL JOB?, Panel consisting of A. Oladko, Plant Mgr., Johnson & Johnson, New Brunswick, N. J., A. V. Feigenbaum, Manager of Programming, Aircraft Nuclear Propulsion Project, General Electric Co., Cincinnati 15, O., H. E. Thompson, Dir. of Res., Anaconda Wire & Cable Co., Hastings on Hudson 6, N. Y., and W. H. Smith, Mgr., Anal. Dept., Quality Control Office, Ford Motor Company, Dearborn, Mich---------- *

B. ARMED SERVICES - ORDNANCE

Moderator - Charles Bicking, Ordnance Engineer, Ordnance Research and Development Division, Executive Branch, Office of the Chief of Ordnance, Washington 25, D. C.

QUALITY CONTROL PROJECT FOR 105MM AMMUNITION, Arthur Stein, Chief, Quality Control Engineering Section, Ordnance Ammunition Center, Joliet, Illinois---------- 455

ARMY ORDNANCE GAGING PROBLEMS AND PRACTICES, M. L. Fruechtenicht, Manager Ordnance Gage Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania---------- 479

C. CHEMICAL INDUSTRY - GENERAL

Moderator - R. S. Bingham, Jr., Quality Control Engineer, Atlas Powder Company, Volunteer Ordnance Works, Chattanooga, Tenn.

STATISTICAL CONTROL IN CONTINUOUS CHEMICAL PROCESSING, Henry W. Isleib, Head, Chemical Extraction Department, Bristol Laboratories, Inc., Syracuse, New York, and Robert M. Hofstead, Special Assistant to Production Manager Bristol Laboratories, Inc., Syracuse, New York---------- 135

EXPERIMENTS IN THE CHEMICAL PLANT, James A. Mitchell, Chief Chemical Engineer, Tennessee Eastman Corporation, Kingsport, Tenn.----- 29

D. METALS INSPECTION

Moderator - Edward M. Schrock, Manager of Quality Control, Koppers Company, Inc., Baltimore 3, Maryland

ERRORS OF MEASUREMENT, H. J. Becker, American Steel & Wire Division, U. S. Steel Corporation, Waukegan 1, Illinois---------- 485

STATISTICAL METHODS FOR PRODUCT WEIGHT CONTROL, A. H. Maack, Quality Control Engineer, and J. R. Brady, Supervisor of Research and Testing, International Harvester Company, Wisconsin Steel Works, Chicago 17, Illinois---------- 267

*Panels are informal. No material available for publication.

E. RELIABILITY OF ELECTRONIC EQUIPMENT

Moderator - M. E. King, Quality Control Division, Bureau of Ordnance, Navy Department, Washington 25, D. C.

APPLICATION OF STATISTICAL METHODS IN EVALUATING PERFORMANCE OF ELECTRONIC EQUIPMENT, G. Ronald Herd, Chief Statistician and Supervisor of the Statistical Department, Military Contact Division, Aeronautical Radio, Inc., Washington, D. C.--------- 309

STATISTICAL APPRAISAL OF VACUUM TUBE RELIABILITY, R. D. Guild, Sylvania Electric Products, Inc., Emporium, Pennsylvania------ 225

F. APPLICATIONS IN THE PROCESS INDUSTRIES

Moderator - E. C. Mayo, Supervisor of Quality Control, American Viscose Corp., Parkersburg, West Virginia

ATTAINING UNIFORMITY IN PROCESSING RUBBER BATCHES, Paul E. Gnaedinger, Director of Quality Control, Dominion Rubber Company, Limited, Montreal, Quebec---------- 55

SOME APPLICATIONS OF STATISTICAL QUALITY CONTROL IN THE PAPER BOXBOARD AND PRINTING INDUSTRIES, B. E. Sooy, Chief of Quality Control, The Gardner Board & Carton Company, Middletown, Ohio- 605

G. PLASTICS MANUFACTURING

Moderator - E. B. Haden, Director of Quality Control, Esterbrook Pen Company, Camden 1, N. J.

STATISTICAL QUALITY CONTROL IN THE PLASTICS INDUSTRY, R. C. Boyd, Head, Thermoplastics Works Laboratory, Bakelite Company Division, Union Carbide & Carbon Corp., Bound Brook, N. J.---- 601

INDUSTRIAL STATISTICS IN PLASTICS EVALUATION, C.H. Adams, Group Leader, Evaluation Research, and E. C. Harrington, Jr., Manager, Industrial Statistics, Monsanto Chemical Company, Springfield, Mass.---------- 321

H. ADVANCED EDUCATIONAL SERIES

Moderator - I. S. Hoffer, Temple University, Philadelphia 22, Pa.

INTRODUCTION TO EDUCATIONAL SESSION AND THE VARIANCE RATIO TEST, E. C. Harrington, Jr., Manager, Industrial Statistics, Monsanto Chemical Company, Springfield, Massachusetts---------- 315

PERIOD II - THURSDAY, MAY 28, 9:30 - 11:30 A.M.

A. GENERAL INTEREST

Moderator - R. S. Inglis, Superintendent, Quality Control and Inspection, General Electric Company, Philadelphia 42, Pa.

A TRADE UNIONIST LOOKS AT QUALITY CONTROL, William Gomberg, Director of the Management Engineering Department, International Ladies' Garment Workers' Union, New York 19, New York---------- 561

HUMAN RELATION ASPECT OF QUALITY CONTROL, Carl E. Hoover, Quality Control Administrator, Perfect Circle Corporation, Hagerstown, Indiana---------- 219

B. ARMED SERVICES - SAMPLING INSPECTION

Moderator - H. R. Pabst, Jr., Chief Statistician, Quality Control Division, Department of the Navy, Bureau of Ordnance, Washington 25, D. C.

ACCEPTANCE INSPECTION INSPECTION BY VARIABLES, CONCEPTS AND PROCEDURE FOR, Harry M. Rosenblatt, Head, Statistics Branch, Quality Control Division, Department of the Navy, Bureau of Ordnance, Washington 25, D. C.---------- 231

CONTINUOUS SAMPLING PLANS, Harry H. Weingarten, Quality Control Division, Bureau of Ordnance, Navy Department, Washington 25, D. C.---------- 5

C. CHEMICAL INDUSTRY - PHARMACEUTICAL APPLICATIONS

Moderator - P. A. Geary, Package Development, Smith, Kline, & French Laboratories, Philadelphia, Pennsylvania

APPLICATIONS OF STATISTICAL QUALITY CONTROL IN THE PHARMACEUTICAL INDUSTRY, T. J. Moran, Statistical Quality Control Supervisor, Wyeth, Inc., West Chester, Pennsylvania---------- 675

THE ROLE OF A BIOMETRICIAN IN THE PHARMACEUTICAL INDUSTRY, C. I Bliss, Biometrician, The Connecticut Agricultural Experiment Station, New Haven 4, Connecticut---------- 659

D. AUTOMOTIVE INDUSTRY

Moderator - Nelson Meagley, Statistical Quality Control Manager, Willys Overland Motors, Inc., Toledo 1, Ohio

MILITARY INSPECTION PROCEDURES OF VEHICLES AND PARTS, J. P. Roche, Chief, Quality Control, Detroit Arsenal, Centerline, Michigan- 179

QUALITY CONTROL IN INTERNATIONAL HARVESTER FOUNDRIES, James M. Barrabee, Inspection Methods Research Engineer, International Harvester Company, Manufacturing Research Department, Chicago, Illinois-- 493

E. LIFE TESTING ELECTRICAL AND ELECTRONIC EQUIPMENT

Moderator - Leon Bass, Quality Control Supervisor, General Electric Company, Electronics Park, Syracuse, New York

STATISTICAL PROBLEMS IN LIFE TESTING, Benjamin Epstein, Associate Professor, Department of Mathematics, Wayne University, Detroit 1, Michigan---------- 385

HOW RELIABLE IS YOUR LIFE TEST PROCEDURE?, J. Alfred Davies, Supervisor, Statistical Control Methods Section, General Electric Company, Owensboro, Kentucky---------- 255

F. MATHEMATICAL STATISTICS - THEORY

Moderator - Mason E. Wescott, Mathematics Department, University College, Rutgers University, New Brunswick, New Jersey

LINEAR PROGRAMMING, A. W. Tucker, Professor of Mathematics, Princeton University, Princeton, New Jersey---------- 651

SOME ASPECTS OF EXPERIMENTAL DESIGN, Lee Crump, Head, Statistics Section, Atomic Energy Project, University of Rochester, Rochester, New York---------- 155

G. QUALITY CONTROL APPLICATIONS IN THE FOOD INDUSTRY

Moderator - Frank Kenney, P. Ballantine & Sons, Newark, New Jersey

THE EFFECT OF THE CONCENTRATION AND FLAVOR INTENSITY IN A FOOD PRODUCT ON THE ORGANOLEPTIC SENSES, D. A. Brandt, E. S. Edelmann, and N. A. Forte, Schenley Distillers, Cincinnati 2, Ohio------ 505

PROCESS CONTROL IN THE BREWING INDUSTRY BY STATISTICAL METHODS, Robert I. Tenney, President, Wahl-Henius Institute, Chicago, Illinois---------- 377

H. ADVANCED EDUCATIONAL SERIES

Moderator - I. S. Hoffer, Temple University, Philadelphia 22, Pa.

SAMPLING VARIATION OF AVERAGES, TESTING OF HYPOTHESES, AND SAMPLE SIZES, Norman R. Garner, Quality Assurance Office, PQMD, General Testing Laboratories, Philadelphia, Pa.--------- 87

PERIOD III - THURSDAY, MAY 28, 2:30-4:30 P.M.

A. QUALITY CONTROL EDUCATION AND TRAINING

Moderator - E. May, Assistant Director of Quality Control, Johnson & Johnson, New Brunswick, New Jersey

COMPANY EDUCATION ON THE QUALITY PROBLEM, A. V. Feigenbaum, Manager of Programming, Aircraft Nuclear Propulsion Project, General Electric Company, Cincinnati 15, Ohio---------- 671

METHODS OF TRAINING FOR QUALITY CONTROL ENGINEERS AND SUPERVISORS, Ralph E. Wareham, Consultant on Quality Control, Chappaqua, New York---------- 331

B. ARMED SERVICES - AIRCRAFT

Moderator - Dorian Shainin, Chief Engineer, Rath & Strong, Inc., Boston 10, Massachusetts

QUALITY CONTROL IN THE AIRCRAFT INDUSTRY - Film
PRACTICAL APPLICATION OF STATISTICAL QUALITY CONTROLS IN THE AIRFRAME INDUSTRY, Phil Harr, Quality Manager, Fairchild Aircraft Division, Fairchild Engine and Airplane Corporation, Hagerstown, Maryland---------- 161

C. CHEMICAL INDUSTRY - PETROLEUM APPLICATIONS

Moderator - B. H. Lloyd, Statistical Analysis Specialist, Canadian Industries, Ltd., Montreal, Quebec, Canada

APPLICATIONS OF STATISTICAL METHODS IN RESEARCH AND DEVELOPMENT, James H. Matheny, The Texas Company, Beacon, New York-------- 677

SOME APPLICATIONS OF STATISTICS IN A REFINERY, C. T. Shewell, Research Specialist, Humble Oil and Refining Company, Baytown, Texas---------- 101

D. COMPUTER APPLICATIONS IN THE METALS INDUSTRY

Moderator - J. V. Sturtevant, Process Metallurgist, United States Steel Corporation, Pittsburgh, Pennsylvania

THE APPLICATION OF AN I.B.M. 604 ELECTRONIC CALCULATING PUNCH TO SOLVE MULTIPLE LINEAR REGRESSION PROBLEMS, D. H. W. Allan, and R. F. Attridge, Processing Division, Metallurgical Department, The Steel Company of Canada, Ltd., Hamilton, Canada---------- 521

APPLICATION OF CONSOLIDATED COMPUTER TO THE SOLUTION OF A METALLURGICAL CORRELATION PROBLEM, Donald S. Leckie, Quality Control Engineer, Republic Steel Corporation, Cleveland, Ohio, and Richard F. Cain, Assistant Superintendent, Quality Control, Republic Steel Corporation, Canton, Ohio---------- 39

E. GAGING AND GAGE DESIGN

Moderator - F. B. Murphy, Field Engineer, Johnson Gauge Company, Bloomfield, Connecticut

GAGE CONTROL, C. H. Borneman, Manufacturing Engineering Services Department, General Electric Company, Schenectady 5, New York- 535

QUALITY CONTROL'S IMPACT ON THE USE OF THE STANDARD INCH, John A. Harrington, Chief Engineer of the DoALL Company, Des Plaines, Illinois---------- 127

F. VENDOR RELATIONS

Moderator - F. C. Schulze, Chief Inspector, Hunter Spring Company, Lansdale, Pennsylvania

WHO CONTROLS THE VENDOR'S QUALITY, J. Philip Worth, Chief Inspector, Electrolux Corporation, Old Greenwich, Connecticut---------- 97

VENDOR RESPONSE TO THE QUALITY CONTROL APPROACH, C. R. Burdick, Supervisor, Statistical Analysis Section, Quality Control Office, Ford Motor Company, Dearborn, Michigan---------- 419

G. INFLUENCES ON PRODUCT QUALITY

Moderator - J. E. Beezer, RCA Moorestown Engineering Plant, Moorestown, New Jersey

SPECIALIZED LIGHTING FOR INSPECTION AND CONTROL, E. A. Linsday, Electrical Engineer, Lamp Division, General Electric Company, Nela Park, Cleveland 12, Ohio---------- 357

PRETESTING PACKAGED PRODUCTS FOR SAFE TRANSIT - A VITAL PHASE OF QUALITY CONTROL, Ralph F. Bisbee, Manager of Quality Control, Westinghouse Electric Corporation, Mansfield, Ohio---------- 365

H. ADVANCED EDUCATIONAL SERIES

Moderator - I. S. Hoffer, Temple University, Philadelphia 22, Pa.

BACKGROUND OF EXPERIMENTAL DESIGN, M. A. Brumbaugh, Director, Department of Statistical Analysis, Bristol Laboratories, Inc., Syracuse, New York---------- 435
(This session and Period IV, session H will constitute an integrated presentation.)

PERIOD IV - FRIDAY, MAY 29, 9:30 - 11:30 A.M.

A. ENGINEERING AND QUALITY CONTROL

Moderator - G. R. Page, Engineer in Charge of Quality Control, Western Electric Company, Inc., Baltimore 24, Maryland

ANALYSIS OF DIMENSIONS AND TOLERANCES, Earle Buckingham, Professor of Mechanical Engineering, Emeritus, Massachusetts Institute of Technology, Cambridge 39, Massachusetts---------- 21

STATISTICAL METHODS AND THE ENGINEER, James H. Davidson, Head of Statistical Methods Section, Electromechanical Division, General Engineering Laboratory, General Electric Company, Schenectady 5, New York---------- 335

B. ARMED SERVICES - MIL - Q - 5923A

Moderator - J. Manuele, Director, Headquarters Quality Control, Westinghouse Electric Corporation, Pittsburgh 21, Pa.

MIL-Q-5923A, Panel consisting of J. G. Schneider, Colonel, U.S.A.F., Acting Chief, Quality Control Division, Headquarters, Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio, O. C. Griffith, Lt. Colonel, U.S.A.F., Chief, Operations Office, Quality Control, Headquarters, Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio, Joseph Paul, Quality Control Manager, Glenn L. Martin Company, Baltimore 3, Maryland, and H. Simpson, General Electric Company, A.G.T. Division, Cincinnati, Ohio---------- *

C. TEXTILE INDUSTRY

Moderator - G. M. Hailes, Quality Control Manager, Avondale Mills, Sylacauga, Alabama

CONTROL CHARTS AND QUALITY LEVEL APPLICATIONS IN TEXTILES, Edward S. Rudnick, Director of Research, Wamsutta Mills, New Bedford, Massachusetts---------- 65

EVENNESS TESTING APPLIED TO TEXTILE QUALITY CONTROL, Panel consisting of Moderator, D. S. Hamby, School of Textiles, North Carolina State, Raleigh, North Carolina, and members, F. J. O'Neil, Director, Physical Research Department, Pacific Mills, Lawrence, Massachusetts, S. E. Patrick, Jr., Research and Development Department, Saco-Lowell Shops, Biddeford, Maine, L. H. Hance, Institute of Textile Technology, Charlottesville, Virginia, and J. W. Stuart, Director of Sales, Uster Corp., Charlotte, North Carolina---------- *

D. METALS INDUSTRY

Moderator - W. P. Goepfert, Head, Statistical Quality Control Department, Aluminum Company of America, Pittsburgh 19, Pa.

QUALITY CONTROL AT THE AMERICAN BRASS COMPANY, F. G. Smith, Assistant Metallurgist, The American Brass Company, Waterbury, 20, Connecticut---------- 1

TESTING PRODUCTION DATA FOR NORMALITY, F. G. Norris, Metallurgical Engineer, Wheeling Steel Corporation, Steubenville, Ohio----- 71

E. QUALITY CONTROL IN PRECISION MANUFACTURING

Moderator - A. B. Mundel, Chief Engineer, Sonotone Corporation, Elmsford, New York

A PRACTICAL APPROACH TO LOSS REDUCTION IN PRECISION MANUFACTURING, Edward R. Close, Chief Inspector, Instrument Division, Bausch & Lomb Optical Company, Rochester, New York---------- 341

CONTROLLING VISUAL CHARACTERISTICS ON PRECISION PARTS, A. S. Hayes, Quality Control Supervisor, Perfect Circle Corp., Hagerstown, Indiana---------- 213

F. OPERATIONS RESEARCH

Moderator - P. S. Olmstead, Bell Telephone Laboratories, Murray Hill, New Jersey

INDUSTRIAL APPLICATIONS OF OPERATIONS RESEARCH, John B. Lathrop, Operations Research Division, Arthur D. Little, Inc., Cambridge, Massachusetts---------- 295

THE PLACE OF OPERATIONS RESEARCH VS. QUALITY CONTROL, Panel consisting of Moderator, P. S. Olmstead, Bell Telephone Laboratories, Murray Hill, New Jersey, and members, S. S. Wilks, Princeton University, Princeton, New Jersey, W. R. Weaver, Republic Steel Corp., Cleveland 1, Ohio, J. B. Lathrop, Arthur D. Little, Inc., Hugh Smallwood, U. S. Rubber Company, Passaic, New Jersey---------- *

G. INDUSTRIAL ENGINEERING

Moderator - E. E. Schiesel, Chief Engineer, Mattatuck Manufacturing Company, Waterbury 20, Massachusetts

INDUSTRIAL ENGINEERING AND QUALITY CONTROL, Arnold Longley, Manager of Training, Atlas Steels Limited, Welland, Ontario, Canada-- 199

WORK SAMPLING (RATIO DELAY STUDY) TECHNIQUE, C. L. Brisley, Industrial Engineering Manager, Wolverine Tube Division, Calumet & Hecla, Inc., Detroit, Michigan---------- 115

H. ADVANCED EDUCATIONAL SERIES

Moderator - I. S. Hoffer, Temple University, Philadelphia 22, Pa.

ON THE DESIGN OF EXPERIMENTS, D. B. De Lury, Director, Department of Mathematical Statistics, Ontario Research Foundation, Toronto, Canada---------- 439
(This session and Period III, session H will constitute an integrated presentation.)

*Panels are informal. No material available for publication.

PERIOD V - FRIDAY, MAY 29, 2:30 - 4:30 P.M.

A. GENERAL INTEREST

Moderator - H. W. Isleib, Head, Chemical Extraction Department, Bristol Laboratories, Inc., Syracuse, New York

INSTALLING A QUALITY CONTROL PROGRAM, H. T. Hallowell, Jr., President, Standard Pressed Steel Comapny, Jenkintown, Pennsylvania----- *

MANAGEMENT'S USES OF QUALITY CONTROL, C. R. Scott, Jr., Manager, Otis Elevator Company, New York 1, New York---------- 643

B. ARMED SERVICES - MIL STANDARDS

Moderator - J. Sidebottom, Assistant Director, Technical Service, Aircraft Industries Assoc., Washington 5, D. C.

MILITARY SPECIFICATIONS AND STANDARDS AS STANDARDIZATION TOOLS - THEIR DEVELOPMENT AND COORDINATION WITH INDUSTRY, Russell A. Moody, Director, Office of Standardization, Defense Supply Management Agency, Washington 25, D. C.---------- 541

C. TEXTILE INDUSTRY

Moderator - W. S. Smith, West Point Manufacturing Company, Research Division, Shawmut, Alabama

QUALITY CONTROL OF VISUAL CHARACTERISTICS IN CARPET MANUFACTURE, Albert G. Klock, Quality Control Superintendent, Bigelow-Sanford Carpet Company, Inc., Thompsonville, Connecticut----- 291

ACCEPTANCE INSPECTION FOR THE CONSUMER, Jules Labarthe, Jr., Senior Fellow, Commodity Standards Fellowship, Mellon Institute of Industrial Research, University of Pittsburgh, Pittsburgh 13, Pennsylvania---------- 205

D. PLANT TOUR, QM General Testing Laboratories, Quartermaster Depot, and Manufacturing Division, Philadelphia, Pa.

E. PLANT TOUR, Ford Motor Company, Chester, Pa.

F. MATHEMATICAL STATISTICS

Moderator - S. S. Wilks, Department of Mathematics, Princeton University, Princeton, New Jersey

SOME BASIC PRINCIPLES OF SAMPLING, W. Edwards Deming, Bureau of the Budget and the Graduate School of Business Administration, New York University---------- 427

CHI-SQUARE TESTS OF INDEPENDENCE - EXPOSITION AND COMMENT, Acheson J. Duncan, Associate Professor of Statistics, The Johns Hopkins University, Baltimore 18, Maryland---------- 167

G. PLANT TOUR, The Budd Company, Philadelphia, Pa.

---

*No material available for publication for this session.

H. ADVANCED EDUCATIONAL SERIES

Moderator - I. S. Hoffer, Temple University, Philadelphia 22, Pa.

THE TECHNIQUE OF REGRESSION ANALYSIS, Miss B. B. Day, Principal Statistician, U. S. Naval Engineering Experimental Station, Annapolis, Maryland, F. R. Del Priore, and E. Sax, Statisticians, U. S. Naval Engineering Experimental Station, Annapolis, Maryland---------- 399

ALTERNATES

NEW DEVELOPMENTS IN AIR GAGING, F. Meyer, Jr., Chief Engineer, Instrument Gage Department, The Taft-Pierce Manufacturing Company, Woonsocket, Rhode Island---------- 471

COST OF QUALITY, W. H. Lesser, Supervisor of Quality Control, Control Department, General Electric Company, Schenectady, New York---------- 551

THE SPAN PLAN - A NEW METHOD OF PROCESS CAPABILITY ANALYSIS, Leonard A. Seder, Chief Qiality Control Engineer, and David Cowan, Research Analyst, Gillette Safety Razor Company, Boston, Massachusetts---------- 567

CONVENTION COMMITTEES
7th ANNUAL A.S.Q.C. CONVENTION

General Chairman
W. J. MASSER, General Electric Company

Secretary-Treasurer
F. R. BAILEY, The Electric Storage Battery Company

Publicity
J. MEHALEK, Chairman, The Budd Company
C. LARSON, Frankford Arsenal
R. JACOBS, RCA Victor Company
H. WOHLERS, The Autocar Company

Registration
J. V. HART, Chairman, Minneapolis-Honeywell Regulator Co., Brown Instrument Division
P. J. RIES, Minneapolis-Honeywell Regulator Co., Brown Instrument Div.
W. W. FRANKHOUSER, Sharpe & Dohme, Inc.

Arrangements
G. R. GEIGER, Chairman, International Resistance Company
W. R. REEVES, RCA Victor
T. J. MORAN, Wyeth, Inc.
J. M. SHERMAN, Standard Pressed Steel

Program
A. R. CRAWFORD, Chairman, Rutgers University
S. S. HART, The Esterbrook Pen Company
J. G. RUTHERFORD, Glenn L. Martin Company

Proceedings
E. M. SCHROCK, Koppers Company, Inc.

Exhibits
S. WIRTSCHAFTER, Chairman, I-T-E Circuit Breaker Company
E. E. PATTERSON, I-T-E Circuit Breaker Company
P. I. McGRATH, JR., U. S. Pipe & Foundry Company
J. BRAVERMAN, Frankford Arsenal

Reception
A. F. BOTTKE, Chairman, Ternstedt Div., General Motors Corp.
C. J. McCLINTOCK, Hunter Spring Company
F. C. SCHULZE, Hunter Spring Company
M. E. WOLCOTT, Arcos Corp.

National Exhibits Chairman
C. R. KENDEL, Lamson & Sessions Company, Cleveland, Ohio

QUALITY CONTROL AT THE AMERICAN BRASS COMPANY

F. G. Smith
The American Brass Company

The first experimental application of Quality Control at the American Brass Company was with the production of hot pressed aluminum bronze valve-seats. The objective was to reduce the number of tests required to insure that hardness was within specified limits. We used $\bar{X}$ and R charts, samples of 5, for consecutive lots of 750 and kept the product under control. The few lots which gave values outside control limits were checked with suitable double sampling system from Dodge & Romig tables. Very seldom was a lot of 750 held up for reprocessing or to be scrapped. The number of tests was reduced to about 1/10 of what it had been, and the product was improved.

The next application of the same kind was on copper content of cartridge brass. A statistical analysis of 6 months previous data showed us that 4 analyses a day of random samples from 40 furnaces running continuously on 8 hour shifts would tell how the copper content was running. For example, during the year 1944, there were 1252 analyses made from about 180,000 heats of metal cast. The average copper content was 69.960%, standard deviation 0.4% vs. specification requirements 70.00% $\pm$ 1.50%, or standard deviation 0.5%. This was not intended as a means of direct control. It was simply a daily record (one day old to be sure) which showed us how copper content was running. The number of analyses was reduced by 2/3. This saved time and labor for other work during a critical period.

About this time in 1944 I attended a War Production Board 7 day course at Purdue. The principal thing learned there was that about half the companies represented by the attendance were American Brass customers. If these concerns were interested, I concluded that The American Brass Company should accept and make general use of Statistical Quality Control.

Since that time, in all our mill laboratories, which are under the direction of the Central Technical Department, there has been continuing and serious attempt to use Quality Control methods. Laboratory personnel have attended many elementary instruction courses, principally those given by local chapters of A.S.Q.C.

The application of Quality Control at our 10 plants varies because of the different kinds of products. In every mill, quality control was accepted with enthusiasm but, in time, experience has shown that the ordinary $\bar{X}$ and R or C charts are not generally applicable to our principal products.

What we have learned is that statistical analysis of data collected in an orderly manner is a means of telling us definitely the practical limitations as to composition, dimensional tolerances, physical properties, and surface characteristics of our products. For example, the physical properties and Rockwell hardness and their standard deviations have been correlated against per cent reduction by rolling for brass and bronze strip metal. With this base, the effects of composition on the physical properties can be evaluated. We can then adjust the rolling temper to maintain a desired range of physical properties.

Statistical analysis has been very useful in the study of thickness of lamp base brass, the eccentricity of tubing, and the diameter of rods. What we have found enables us to standardize, improve, or even change our production methods to obtain satisfactory results.

We have a Direct Reading Spectrometer for the rapid analysis of copper alloys at one of our mills. It would not have been possible to evaluate results and make necessary adjustments and calibrations without the use of statistical analysis. Some of the problems have required analysis by multiple correlation, and a study of interaction between variables.

One class of tubes drawn to stock size in one mill and finished in another are very carefully inspected at the stock sizes for physical imperfections. A double sampling acceptance inspection per Dodge & Romig, Table Da-7 (Average AOQL 7%), has served the purpose of keeping the quality satisfactory, and there are no more arguments about it based on opinions.

Another mill uses Dodge & Romig Table SL-5 (Process Average 0-.05, Lot Tolerance 5%, AOQL about 1%) for shipments of tubes to particular customers who process the tubes further.

Hot pressed forgings at one plant are inspected before shipment using Dodge & Romig single sampling Table SL-2 (Process Average 0.5%, Lot Tolerance 2%). This saves inspection time because in the past all lots were inspected 100%.

Another application deals with process control where we specify the weight and weight tolerances for billets to be made into tubes and into rods where the finished product is cut to length. This insures the maximum number of lengths finished from a billet, with control of cutting scrap at as low an amount as possible. This sort of control has always been a production problem in a brass mill where scrap made in running down or cutting to length is an expensive item. Statistical analysis tells us more accurately than heretofore what must be done to keep such process scrap at the lowest level.

A similar application is in determining the weight of mill stock tube sizes to enter upon a specific order so that the finished weight will not be less than that ordered nor very much more. If the weight is less, a special small lot has to be made to complete the order, or if it is more than 5% greater, that extra amount cannot be shipped. The excess has to be stocked or possibly finished to another smaller size. Either over- or under-weight must be avoided as much as possible.

The foregoing outlines a few general applications which are typical of what we are doing, but there is one which I think will be interesting if outlined in detail. That is quality control as applied to the production of 30 and 50 caliber cartridge case cups and bullet jacket cups. Cartridge case cups are made of 70-30 brass; bullet jackets of 90-10 Gilding.

Composition

Composition is determined at one mill on a sample from each furnace and each shift every day. Average and range charts are used for control. Trends in copper content, for example, are adjusted by changing the zinc weighed into the furnace charges. Impurities, such as lead and iron, present no particular problem.

Gauge of Strip Metal

The gauge of the strip and the diameter of the blank (both of which affect the cup dimension) are determined from a correlation against the specified grain weight of the cup. This saves metal and stops heated argument between rolling mill and press room foremen.

Sampling of Cups for Dimensions

Cups, as they fall from each punch and die, accumulate in separate containers identified by Hole Number. About every 20 minutes, a cup is taken as it falls from the press at each hold and over a 4 hour period, during each shift, until a sample of 12 or 15 cups per hold is collected every 4 hours. From these samples 2 cups are taken at random and gauged for outside diameter, wall thickness, wall variation and base thickness. The readings on each cup are charted in reference to control lines at the specification tolerance and at a press control tolerance. The latter is determined from a careful study of the respective standard deviations for such consecutive samples. Lines are drawn on this basis, well within the specification tolerance. In this way, trends due to tool wear are shown. If any dimension falls on or just outside the press control limits, the tool setter makes the necessary corrective adjustment. If any dimension falls close to or outside the specified tolerance, all the sample cups from that hole are gauged. If any of these extra readings are at the specified tolerance or outside, the barrel into which all the cups from that hold have been collected and tagged with hole number are tagged again with red and checked. A random sample of 100 cups is taken from this barrel. If 10 minor or 2 major dimensional defects are found, the barrel lot is scrapped. A minor dimensional defect is one within a very small tolerance allowed outside the dimension specified. A major dimensional defect is one outside this tolerance.

The wall height variation and mouth bevel present no serious problem and do not require following with control charts. They are taken care of in the final general acceptance inspection by MIL-STD-105A.

Grain Size

The cups go from the press room to a continuous annealing furnace where, for example, about 75,000 cups are annealed per hour. The furnace operator takes 2 cups every 15 minutes and obtains the average Rockwell F hardness on each. The mill laboratory takes five (5) cups at the furnace every hour, obtains the average Rockwell F hardness on each, selects cups with high, low and middle hardness for grain size readings. Grain size is continuously correlated vs. Rockwell hardness. From this correlation, hardness limitations are established for the furnace operator. It is his responsibility to control his furnace according to his own tests, made every 15 minutes. The laboratory is responsible for the grain size control through its guidance of the furnace operator. This procedure is a good example of what can be done with a very small sample from a continuous process, when that process is very carefully controlled.

General Inspection

The finished cups are inspected for dimensional and visual defects following MIL-STD-105A, Table IV-B, double sampling, acceptable qualitative level 0.25%. Lot size Q, over 550,000. First sample 1000, accept at 5, reject at 17. Second sample 2000, accept at 16, reject at 17.

We take 15 random cups from each lot for grain size and make affidavit of results which is accepted by the inspector.

Finally the Government inspects all the lots again following MIL-STD-105A, Table IV-B.

This Quality Control Procedure has made it possible to meet economically the present requirements of MIL-C-10375 for case cups and MIL-C-3383 for jacket cups.

CONTINUOUS SAMPLING PLANS

Harry Weingarten
Quality Control Division
Bureau of Ordnance
Navy Department

Introduction. The Quality Control Division of the Bureau of Ordnance, Department of the Navy, is charged with many responsibilities revolving around the task of supplying the Fleet with ordnance of all types. One of the main responsibilities deals with the acceptance inspection program applied to material manufactured for the Bureau by private contractors. For many years and still at present, MilStd 105A "Sampling Procedures and Tables for Inspection by Attributes" has been the basis on which this program rests. Briefly, MilStd 105A can be characterized by saying that a decision to accept or reject a lot is made on the basis of a random sample from the lot. A sample selected at random from a lot is inspected and each item in the sample is classified into one of two categories, defective or non-defective. A figure is set so that if the number of defective items found in the sample is equal to or larger than it, the decision is to reject the lot. Otherwise, the lot is accepted.

In this type of sampling, called single sampling, the two quantities, the size of the sample n, and the rejection number r completely determine the effect of lot by lot use of the sampling plans in MilStd 105A. That is to say, for a particular sampling plan the values of n and r determine the frequency with which decisions to accept or reject a lot will be made. Thus, one plan may be preferred to another since for a fixed quality level in the lot, the one plan will more often accept that lot than the other.

Now for various reasons it has been found difficult, undesirable or actually impossible in some cases to apply MilStd 105A to an inspection operation on a moving line. Much of modern production makes use of the moving belt or line and use of MilStd 105A for inspection of material produced in this way has caused objections of the following sort to be raised.

a. Interference with efficient production operations by requiring lotting on the line.

b. Hazards to safety resulting from lotting some explosive material.

c. Disassembly, sometimes destructive, in the case of lot rejections for hidden defects.

d. Insufficient storage facilities for lotting.

e. Possibility of rejecting product not yet produced, e.g., when rejection number was reached but present inspection lot size was not yet produced.

f. Establishing small lot sizes to minimize the foregoing objections - the consequence of this was to increase relative total amount of inspection for all units produced and to decrease the

purchaser's protection because of small sample size per lot.

g. Loss of protection to the purchaser when more than one inspection station was necessary to inspect a unit of product and the sampling plan intended for the unit of product was applied separately to lots accumulated at each inspection station.

These and other difficulties accompanying the system often caused relations between producer and purchaser to be less cordial than they should be and increased the cost of inspection and production.

Let us consider the objections in some detail, remembering that MilStd 105A deals with making decisions about a lot. The first objection arises in situations where there are inspection stations on a line following different production operations. It is desirable to inspect before permitting material to move on to further processing. Requiring the formation of a lot on the line can be seen to entail difficulties, particularly if the lot is of an appreciable size. With regard to (b) it is simply enough to say that it is often dangerous to put a great deal of explosive material in one place at one time. Objection (c) harks back to (a). If inspection is delayed until an item is completed, inspection may require the destruction of the item to uncover the characteristics to be inspected. Objection (e) is perhaps one of some statistical interest. If the lot size were to be produced during a four hour period, the inspector would select his sample from the production during these four hours. It might be possible, and indeed it often occurred, that before the full sample was inspected, and also therefore the lot was not yet produced, the rejection number was reached. At this time, the manner in which to proceed is not clear. To maintain the protection afforded by use of the sampling plans in MilStd 105A, it would be necessary to allow the lot to be produced and then, of course, reject it. It would not be necessary to inspect any further. This seemed wasteful and unfair to the contractor however. On the other hand, to consider that only what had been produced is to be rejected is not fair to the Navy, since for that lot size a smaller rejection number than the one used would have obtained. Clearly MilStd 105A is not satisfactory in such a case and the attempt to rectify the matter as outlined in (f) was not desirable. From the Navy's point of view, objection (g) is a strong one since it might be possible, when 1% defective material is the worst that is desired, to receive as high as 3% defective material if 3 inspection stations were used and no rejection occurred at any station.

Continuous Sampling. OSTD 81, "Sampling Procedures and Tables for Inspection by Attributes on a Moving Line" was designed to meet the Bureau's problems of inspection on a moving belt and overcome the objections raised above. Below are some of the features of the plans, which we may notice before discussing their actual operation.

The plans make

a. Provision for control of the percent defective of product accepted by assuring an average outgoing quality limit (AOQL).

b. Provision for criteria for stopping inspection when the percent defective of product submitted is greater than the acceptable quality level (AQL), and require that the producer be advised of defectives found.

c. Provision for simultaneous application to more than one defect classification, e.g., critical, major A, major B, minor A.

d. Provision for application to the case where inspection of a unit of product requires more than one inspection station and effective purchaser protection for the unit of product.

e. The plans also require the acceptance of all product passing the acceptance inspection station, or the last acceptance inspection station if multiple station inspection is in effect.

f. Require corrective action by the producer for his production process if it is necessary to stop acceptance inspection because of defective product.

g. Require corrective action by the producer for units produced and on the line between point of production and last point of acceptance inspection.

h. Require that the acceptance inspector use his experience and knowledge of the production process to assure himself that effective corrective action has been taken by the producer before he may resume inspection after a forced stop.

i. Require special Bureau attention for action if the acceptance inspector reports unsatisfactory corrective measures by the producer, or if defective product continues to be submitted after tightened inspection is in effect.

Let us now turn to the actual operation of a plan in the simplest possible situation. We will assume that there is only one point of inspection for an item and that the item is inspected for only one defect classification. The inspectors job can now be outlined as follows: (For reference we include here Table 2 taken from OSTD 81).

A symbolic representation of the production point and inspection station on a production line is of use in the application of this plan.

P represents the end of a manufacturing process before a point of inspection I. The items flow along the line to the inspector located at I, and from then on to further processing, storing or packaging.

a. The inspector begins his job by estimating the amount of production to pass the point of inspection on the line during the day (or shift).

b. He refers to Table 2 and in the row corresponding to this estimate together with the column corresponding to the defect classification under consideration, he finds four numbers, one in the column headed f, one in the column headed i, and two in the column headed a. The quantities f, i, and a are to be used in his inspection procedure. At the start of inspection on the contract, the a for normal inspection is used.

c. The inspector is now ready to begin inspection. He starts first with consecutive item inspection, that is, inspection of every

Days Production (Items)	Sampling Rate	DEFECT CLASSIFICATION								
		Major A			Major or Major B			Minor A		
			a			a			a	
	f	i	Normal Insp.	Tightened Insp.	i	Normal Insp.	Tightened Insp.	i	Normal Insp.	Tightened Insp.
66 - 110	1/10	40	0	0	20	0	0	10	0	0
111 - 180	1/10	40	0	0	20	0	0	10	1	0
181 - 300	1/10	40	0	0	20	1	0	10	2	1
301 - 500	1/10	40	0	0	20	1	0	10	3	2
501 - 800	1/10	40	0	0	20	2	1	10	4	3
801 - 1300	1/10	40	0	0	20	3	2	10	6	4
1301 - 3200	1/15	50	1	0	25	4	3	10	8	6
3201 - 8000	1/25	60	2	1	30	5	4	15	11	8
8001 - 22000	1/50	80	3	2	40	7	5	20	14	11

Table 2

i = required length of sequence of consecutive items to be inspected and found free of defects before sampling inspection at the rate f can begin.

f = sampling rate (for example 1/10 means 1 in 10).

a = maximum allowable number of sequences of consecutive item inspection beyond the first without requiring interruption of inspection; a is also the maximum allowable number of defectives.

item consecutively, until a sequence of i consecutive items are found free of defectives.

d. When this has occurred, he begins sampling at the rate indicated by the number f. For example, if the f being used in a particular situation is 1/10, this means that during the time the inspector is choosing items at random for sampling, he will choose one item for every 10 that have gone by. This does not mean that he must conscientiously choose one in every consecutive 10 produced. He may, for example, for a flow of 20 items, pick 2 from the second or first 10. For a sequence of 100 items he may choose his 10 in any way that is not accompanied by a definite rule. Thus the 10 in the 100 might be the 1st, 8th, 19th, 21st, 43rd, 48th, 67th, 75th, 91st and 95th. The objective in not following a definite rule, is to prevent anyone from learning this rule and introducing perfect product into the line at the appropriate sampling points.

e. If a defective is found during the sampling at the rate f, the inspector will revert immediately to consecutive item inspection until again i consecutive items are found free of defectives.

f. It is to be understood that the occurrence of a defective counts as a requirement for consecutive item inspection. This means that during the consecutive item inspection period, if a defective is found before i units consecutively are free of defectives, it is to be considered that one consecutive item inspection period is finished and another is begun from the point where the defective is found.

g. If only a or fewer consecutive item inspections occur beyond the first during the day, there will be no interruption of inspection, and all the product that has passed the inspector is accepted.

h. As soon as the a + 1 st consecutive item inspection is required, the inspector is to stop his inspection. At this time, all the product on the line from the point P which has not reached the inspector is not accepted. In all cases, any product that has passed the inspector is accepted.

i. When the a + 1 st consecutive item inspection is required, the inspector also informs the manufacturer as to which defect occurred and that he will no longer inspect product until the manufacturer locates the source of difficulty and assures him that he has removed the cause for defective product.

j. When the inspector is satisfied with this assurance, he will resume inspection. At this time, he must estimate the remaining production for the day. He will then refer to Table 2 for his new inspection plan, this time using the a for tightened inspection. For example, if, for Major B defects, at the beginning of the day the estimate of production was 3000 items, the inspector was using f = 1/15, i = 25 and the a for normal inspection was 4. If the interruption of inspection took place so that only 600 more items could be produced that day from the time production is resumed, the inspector would then use f = 1/10, i = 20 and the a for tightened inspection equal to 1.

k. At the beginning of the next day, the inspector will proceed as in step a, but this time using the a for tightened inspection. He will use this value of a for at least one full day's production. If

during the time that a tightened plan is in effect, it is again necessary to stop inspection, the procedure of step (j) is followed.

l. Only after a full day of production has passed, with no interruptions of inspection, can the inspector at the next day use the a for normal inspection again.

m. All defective product found during inspection is to be replaced with good product.

The plans for critical defects is basically the same. In this case, however, the number of defectives permitted before the interruption of inspection is zero, and so the instructions for the inspector are somewhat altered. They appear here in full, however, with your indulgence. Table 1 from OSTD 81 is also included for reference.

Table 1

Days Production (Items)	Sampling Rate f	Number in Consecutive Item Inspection Sequence i	
		Normal Inspection	Tightened Inspection
66 - 110	1/1	Total Production	
111 - 180	1/10	80	160
181 - 300	1/10	80	160
301 - 500	1/10	80	160
501 - 800	1/10	80	160
801 - 1300	1/10	80	160
1301 - 3200	1/15	100	200
3201 - 8000	1/25	120	240
8001 - 22000	1/50	160	320

Allowable number of defects, a = 0

a. The inspector begins his job by estimating the amount of production to pass the point of inspection on the line during the day (or shift).

b. He refers to Table 1, and in the row corresponding to this estimate he finds three numbers, one in the column headed f and two in the column headed i. The quantities f and i are to be used in his inspection procedure. At the start of inspection on the contract the i for normal inspection is used.

c. The inspector is now ready to begin the inspection. He starts first with consecutive item inspection, that is, inspection of every

item consecutively, until a sequence of i consecutive items are found free of defectives.

d. When this has occurred, he begins sampling at the rate indicated by the number f. For example, if the f being used in a particular situation is 1/10, this means that during the time the inspector is choosing items at random for sampling, he will choose one item for every 10 that have gone by. This does not mean that he must conscientiously choose one in every consecutive 10 produced. He may, for example, for a flow of 20 items, pick 2 from the second or first 10. For a sequence of 100 items he may choose his 10 in any way that is not accompanied by a definite rule. Thus the 10 in the 100 might be the 1st, 8th, 19th, 21st, 43rd, 48th, 67th, 75th, 91st and 95th. The objective in not following a definite rule, is to prevent anyone learning this rule and introducing perfect product into the line at the appropriate sampling points.

If the estimate of the day's production falls in the first row, all the items produced during the day are to be inspected.

e. If a defective occurs, either during the period of consecutive item inspection or sampling, the inspector is to stop his inspection. At this time, all the product on the line, from the point P, which has not reached the inspector, is not accepted. In all cases, any product that has passed by the inspector is accepted.

f. The inspector also informs the manufacturer as to which defectives occurred and that he will not begin to reinspect product until the manufacturer locates the source of difficulty and assures him that he has removed the cause for defective product.

g. When the inspector is satisfied with this assurance, he will resume inspection. At this time, he must estimate the remaining production for the day. He will then refer to Table 1 for a new inspection plan, this time using the i for tightened inspection. For example, if at the beginning of the day an estimate of production was 3000 items, the inspector was using f = 1/15, i = 100. If the interruption of production took place so that only 600 more items could be produced that day from the time inspection is resumed, the inspector would then use f = 1/10, i = 160. If the estimate of the remainder of the day's production falls in the first row, <u>all</u> the remaining items are to be inspected.

h. At the beginning of the next day the inspector will proceed as in step a, but this time he will use the i for tightened inspection. If during the time tightened inspection is being used, it is again necessary to stop inspection, the procedure of steps f and g is to be followed.

i. Only after a full day of production has passed, with no interruptions of inspection, can the inspector at the next day use the i for normal inspection again.

j. All defective product found is to be replaced by good product.

<u>Combined Use for More Than One Defect Classification</u>. Now the instructions accompanying Table 2 can be used with any of the three defect classes in Table 2 and so altogether there is provision for

inspection of 4 defect classes. The type of inspection problem which may arise must be one of the four following:

1. One inspection station, one defect class.

2. One inspection station, more than one defect class.

3. More than one inspection station, one defect class.

4. More than one inspection station, more than one defect class.

The first case we have already discussed, and to handle the remaining three the inspector may proceed as outlined below: (Note that the sampling rate is the same for each defect classification to facilitate combined use.)

1. Stop consecutive item inspection and begin sampling inspection for each defect classification as soon as the required sequence of i units corresponding to each defect classification is found free of defectives. Thus at the beginning, sampling may be in effect for the lower defect classifications (major B and minor A) and consecutive item inspection for the higher defect classifications (major A and critical).

2. When sampling is in effect for more than one defect classification, use the same sampled item for inspection for all defect classifications being considered.

3. The requirement to begin a sequence of consecutive item inspection (after the first at the start of the day) applies only under the conditions in the following subparagraphs (a) and (b).

a. A defective is found during a sequence of consecutive item inspection which began at the start of the day's inspection, or was initiated because of the following subparagraph (b).

b. A defective is found on an item selected for inspection at the regular sampling rate f.

c. Do not begin a sequence of consecutive item inspection if a defective is observed on an item which is not a "regular item for inspection" in accordance with the above subparagraphs (a) and (b).

d. The sequence of consecutive item inspection, when required, is to be made only for the defect class to which the observed defective belongs.

e. The observance of a critical defective at any time during inspection whether a "regular item for inspection" or not, shall be cause for stopping inspection and following the procedures indicated by paragraph (e) of the continuous sampling plan for critical defects.

4. When it is necessary to stop inspection for any class of defects, inspection is automatically stopped for all other classes of defects which pass the same inspection point.

5. Upon resumption of inspection after a forced stop

a. Follow procedure for tightened inspection (paragraph (g) of

the plan for criticals, or paragraph (j) of the plans for the major or minor class) only for the defect class for which inspection was stopped.

b. For other defect classifications, continue with the same sampling plan if production is resumed the same day it is stopped. Accumulate the results of inspection until the end of the day.

Multistation Inspection. If it has been specified that inspection for a unit of product is to occur at more than one inspection station, the inspector is directed to follow the procedure given below for each defect class under consideration.

1. During the time required for consecutive item inspection it may be necessary to have more than one inspector on the line.

2. However, if because of the speed of the moving production line, and the overall work load, it would be difficult for one inspector to follow the directions given in the inspection procedures related to sampling of one item at a time, the inspector is given permission to select more than one item, up to a maximum of five items for inspection at one time.

3. It is emphasized that the above permission is to be used only when absolutely necessary.

4. All the items chosen at one time at any inspection station, must be inspected at once for the defect classifications under consideration.

5. A cumulative total for all stations is to be kept of the defectives found at each station for each defect classification involved.

6. When the cumulative total exceeds by one the specified number "a", of maximum allowable number of consecutive item inspections beyond the first, inspection is to stop and the procedures outlined in the plans are to be followed. For example:

If a = 4 and the inspector has found
1 defective at Station I
3 defectives at Station II
1 defective at Station III
5 Total defectives

the finding of the fifth defective directs him to stop his inspection, as provided in the body of the plans.

7. If it is necessary to stop inspection, any product which has passed the last inspection station is accepted. Product on the line from the first point (P) up to the last inspection station is not accepted.

Nature of the Continuous Plans. You will have noticed in the above discussion that the concept of a lot does not appear in the plans in OSTD 81. In this way the difficulties due to the lot concept of MilStd 105A are done away with. On the other hand, these plans bear a strong relation to MilStd 105A. What has been done here is simple to describe.

In the 1943 issue of the Annals of Mathematical Statistics, H.F. Dodge published a paper entitled "A Sampling Inspection Plan for Continuous Production". This plan is the basis of the procedure of OSTD 81. The concept of beginning with consecutive item inspection, changing to sampling with the indication of good quality, and then returning to consecutive item inspection upon the finding of a defective, is Dodge's. To this procedure the rejection number feature of MilStd 105A has been added so that the inspection job would not continue indefinitely. Clearly the Navy inspector would be required to do too much inspection if inferior product were submitted by the manufacturer. This was the feature of the Dodge plan which made it unacceptable for use without change. Limiting the amount of consecutive item inspection the Navy inspector is permitted to do, was felt to be the idea which would bring the Dodge plan in line with the concept of providing for contractor incentive for good product and also control the amount of purchaser inspection. Rejecting a lot does this when MilStd 105A is in use, and interrupting inspection, we feel provides this incentive when OSTD 81 is in use. It might also be said that rather than accept or reject lots, the continuous sampling plans accept or reject the production process.

Very briefly, the effect of the addition made to the Dodge plan resulted in a mathematical model which generalizes the original procedure, and introduces problems of great difficulty in order to describe the effect of the plans use. These problems have been substantially solved and operating characteristic curves for the plans have been developed. Characteristically, these OC Curves give a somewhat different answer to what takes place when product of a given quality is submitted. The OC Curve of a MilStd 105A plan gives the probability that a lot will be accepted for various values of the incoming percent defective. However, the OC Curve of an OSTD 81 plan gives the probability that production will not be interrupted for various values of the incoming percent defective. This might also be interpreted as giving the percentage of the time production will not be stopped for values of the incoming percent defective.

Some experience with the practical application of the continuous sampling plans has been gained, and all indications up to the present are favorable. Contrary to expectation, the inspector workload involved has not increased over the MilStd 105A requirement when the day's production is considered to be the lot size for instance. In actual practice it has been found that the plans operate to interrupt inspection very quickly when the quality deteriorates beyond the desired level, and on the other hand, when submitted product is good, few consecutive item inspection sequences beyond the first are required. It should be made clear that no matter what quality is submitted for inspection the plans operate to guarantee an AOQL as noted above, so that accepted product is known to be no worse than some preset limit.

Examples. In conclusion let us consider some hypothetical examples of the use of the plans for inspection jobs. Below will be found four reports giving the following information. Sheet 1 shows the results of inspection at one station for one defect category, namely critical. It will be seen that on the first day of inspection no defectives were found and inspection was not interrupted. Thus 80 items were inspected during the initial period of consecutive item inspection and in the subsequent sampling 63 items were dealt with. Since a sampling rate of 1/10 was in effect, this means that in total 80 + 10 (63) = 710 (approximately) items passed the point of inspection and were thus

accepted. On 9/2/52 the initial period of inspection was passed, but the 25th item inspected during sampling was a defective. Upon the resumption of inspection a tightened plan was put into effect requiring 160 consecutive items free of defects before sampling could begin. Notice when no sampling is reported on a line, this indicates that the consecutive item requirement was not met.

Sheet 2 presents a similar report for a major B defect class. Notice that on 9/2/52 the occurrence of the third defective item was what interrupted inspection. In other words, the initial 20 were consecutively free of defectives, the 10th sampling item was defective, the next 20 were consecutively free of defectives but the 25th sampling item was subsequently a defective. Again 20 were consecutively free but the fifth sampling item was a defective. This would require a third period of consecutive inspection beyond the first. The plan allows for only 2 consecutive item inspection sequences beyond the first. Hence inspection is halted and a tightened plan is in effect upon resuming inspection. Notice again in the second line of the report for 9/3/52 no sampling indicates that the defective was found during the consecutive item inspection sequence, and in this case was the tenth consecutive item inspected. Sheets 3 and 4 represent the same inspection operation, one station but now two defect classes. Sheet 3 is a rough work sheet made during the inspection operation itself and Sheet 4 summarizes the information. For process average computation, the formula number of defectives found divided by number inspected is used.

Conclusion. As a final remark, I wish to make one last point clear. The Bureau of Ordnance indicates that the plans in OSTD 81 are mandatory for use when applicable, but it is necessary to assure the Bureau that the inspection operation to which it is desired to apply the plans cannot be handled by MilStd 105A. Indiscriminate use of the plans is not desirable and is to be avoided. On the other hand, we are eager to apply the plans to learn more about their operation in practice, for it is in this way that difficulties in the application will be discovered.

We will be happy to receive comments from anyone interested in these plans since good relations between the Bureau and the contractor is of major importance to us following, naturally, our first concern of protecting the Bureau's interests.

CRITICAL - Individual Use

Sheet 1

Date	Estimated Days Production	Identification				No. of items inspected		No. of Defectives	Start / Stop	Estimate of total accepted/sequence
		f	i	a	N/T	conc.	samp.			
9-1-52	501 - 800	1/10	80	0	N	80	63	0	0800/1600	710
9-2-52	501 - 800	1/10	80	0	N	80	25	1	0800/1130	330
	301 - 500	1/10	160	0	T	160	17	0	1200/1600	330
9-3-52	501 - 800	1/10	160	0	T	160	54	0	0800/1600	700
9-4-52	501 - 800	1/10	80	0	N	60		1	0800/0830	60
	501 - 800	1/10	160	0	T	160	20	1	0835/1100	360
	301 - 500	1/10	160	0	T	160	2	1	1130/1330	180
	111 - 180	1/10	160	0	T	100		0	1430/1600	100
9-5-52	501 - 800	1/10	160	0	T	160	54	0	0800/1600	700
9-6-52	501 - 800	1/10	80	0	N	---	etc.			

MAJOR B - Individual Use

Sheet 2

Date	Estimated Days Production	Identification				Sequence No.	No. of items Inspected		No. of Defect-ives	Start / Stop	Estimate of total accepted/sequence
		f	i	a	N/T		conc.	samp.			
9-1-52	501 - 800	1/10	20	2	N	In	20	68	0	0800/1600	700
9-2-52	501 - 800	1/10	20	2	N	In	20	10	1	0830/	120
						1	20	25	1		270
						2	20	5	1	/1330	70
	181 - 300	1/10	20	0	T	In	20	8	1	1400/1455	100
	111 - 180	1/10	20	0	T	In	20	12	0	1500/1600	140
9-3-52	501 - 800	1/10	20	1	T	In	20	15	1	0800/	170
						1	10		1	/1000	10
	501 - 800	1/10	20	1	T	In	20	50	0	1030/1600	520
9-4-52	501 - 800	1/10	20	1	T	In	20	68	0	0800/1600	700
9-5-52	501 - 800	1/10	20	2	N	In	20	20	1	0800/1100	220
						1	20	46	0	/1600	480
9-6-52	501 - 800	1/10	20	2	N	---	etc.				

Date _____ WORK SHEET Sheet 3

Final form recommended

Defect Classification: CRITICAL							Defect Classification: MAJOR B						
Estimated Production Consecutive	f	i	a	N or T Sampling	No. of Defect-ives	Time inspection Starts Stops	Estimated Production Consecutive	f	i	a	N or T Sampling	No. of Defect-ives	Time inspection Starts & Stops
400	1/10	80	0	N		0800 -	400	1/10	20	1	N		0800
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~							~~1111~~ ~~1111~~						
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~							~~1111~~ 11					1	
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~													
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~							~~1111~~ ~~1111~~						
			1111		1	1030	11					1	1030
300	1/10	160	0	T		1045	400	1/10	20	0	T		1045
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~							~~1111~~ ~~1111~~						1100
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~							~~1111~~ ~~1111~~						
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~									~~1111~~	~~1111~~		1	1130
~~1111~~ 111					1	1200							
							180	1/10	20	0	T		1200
200	1/10	160	0	T		1245	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~						
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~									~~1111~~ ~~1111~~ ~~1111~~	~~1111~~ ~~1111~~			
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~									~~1111~~				1630
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~													
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~													
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~													
~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~													
			1111										

Sheet 4

SUMMARY FORM OF SUGGESTED WORK SHEET IN COMBINED USE

Inspection of 400 items
Illustrative example to demonstrate plan
Oct. 1, 1952 mjg

	CRITICAL 1% Defective	MAJOR B 3% Defective
No. inspected to 1st stop	85	29
Defective found	1	2
% Defective	1.2%	6.9%
No. inspected to 2nd stop	68	29
Defective found	1	1
% Defective	1.5%	3.4%
No. inspected to 3rd stop	13	
Defective found	1	
% Defective	7.7%	
No. inspected after last stop until end of the day	164	45
Defective found	0	0
% Defective	0%	0%
Total No. inspected for day	330	103
Defectives found for day	3	3
% Defective for day	.9%	2.9%
Estimated items accepted for the day	408	406

ANALYSIS OF DIMENSIONS AND TOLERANCES.

Earle Buckingham
Massachusetts Institute of Technology.

The dimensions and tolerances given on the detail drawings of the component parts of a product directly control the tool design, the production set-up, the production and the inspection or quality control procedures. Errors, omissions and over-severe specifications here increase the cost of production enormously.

Efforts are constantly made to reduce troublesome details of every nature to routine. This is one of the main objectives of standardization. At the present time, at least two standardization committees are trying to formulate standards for drafting room procedures, including the problem of dimensioning and the expression of tolerances. Much good can be accomplished along these lines, but in the final analysis, the unique conditions which must be met on each specific product dictate the specific dimensions and tolerances needed by that product. Fortunately the majority of the dimensions for any product are not critical, and these can follow some simple standard procedure. But for the important minority of them which control essential and critical functional conditions, no standard tables of fits, dimensions and tolerances will be satisfactory. The dimensions and tolerances used here must be those that meet the actual functional requirements of the particular product.

There is much that we do not know at the time we start the initial manufacture of a new product. Only the product itself can give us this missing information, and we must definitely seek for it. Some changes in the dimensional specifications are inevitable after the initial production has started. Many of these changes consist of necessary changes in the extent of the tolerances. This is not a matter of personal opinion, or a question of what tolerances can be maintained by the use of a given process, or of the most convenient way to express the dimensions and tolerances to simplify the tooling or the production or the inspection. It is a matter of what the product needs. The ultimate source of information of this nature is in the behavior of the product in the hands of the customers.

Standards can set up the most effective way to express dimensional information, establish a common language of drawings and even standardize conditions of fit for many of the less critical of the operating conditions. But there will always be a few essential and critical conditions on any product where the correct answer can come only from the product itself.

The experimental engineering department is not in the best position to determine and specify the detailed manufacturing requirements for mass production. It can and should establish the initial performance characteristics of the finished product, but this is all. The major interest and effort of this department is to develop a functional design that will accomplish some desired result. Consideration of future manufacturing and control problems here will always be ineffective and tend to distract their thoughts from their major responsibility, the completion of a mechanism that will work.

There are two distinct phases to the design of any product. These

may be called the Inspirational and the Analytical. The experimental or functional design is largely the inspirational with enough of the analytical to make it possible to construct the experimental model.

For mass production, it is essential to make a definite production design. This must be made by engineers familiar with the many detailed problems of mass production. If such a redesign is not made at the start, it will eventually develop itself over a long period of time by trial and error in the course of continued production. This last procedure is a time-wasting and costly method that entails many changes and delays in production.

Production design should be a definite first step in the preparation for the mass production of a new or changed product. The effort here is to achieve a better product that is cheaper to manufacture. It consists of two steps. The first is a careful study of the experimental design to see how it can be made to perform more effectively. The second is a critical study of every feature of every component part of the assembly to see how they can be arranged so that they can be machined most effectively and economically. This is a creative task of a high order. Either the part must be changed to fit the available manufacturing processes and equipment or else some new and improved processes and machines must be developed to fit the requirements of the individual components. The final answer may be a combination of both; most of the features are rearranged to fit the existing facilities, yet a few of them may require definite changes and improvements in the manufacturing processes.

Production design is largely analytical design with enough seasoning of the inspirational to exploit the full functional possibilities of the original experimental design. It is a distinct and separate task from the experimental design. It needs the services of a production design group which is entirely distinct and separate from the experimental design department. The experimental design department should have the veto power over any proposed changes only in case some functional requirements would not be met. Such vetoes should not be sustained if they rest only on personal opinions. The matter should be proved by actual trials.

The results of the production design must next be recorded on the detail drawings of the components and in any other supplementary specifications that may be necessary, so that the components can be made correctly and economically from this information alone. One of the most perplexing problems in attaining this end is that of the dimensional specifications. These dimensional specifications must be so expressed that they have exactly the same meaning to all persons who must use them,- the tool and gage designers, the production set-up and maintenance men, inspectors and others.

Mass production is quite different from machine work in a model or experimental machine shop. The essence of mass production is the breaking down of the complex process of making the whole completed unit by a small group of skilled mechanics into many simple and elementary tasks that can be adequately performed by workers, each with a very limited range of skills. Most of the knowledge, skill and technique formerly brought to the work by the skilled mechanics must be replaced by the information on the detail drawings, while much of the skill and technique must be built into the special equipment which has been

provided. From one to one-hundred or more individual workmen may be called upon, each to contribute their small share towards the completion of a single component part of the product. All of the information they have about the product is derived from the dimensional specifications given on the detail drawing.

It is an axiom in the machine shop that the accuracy to which we can work depends upon the accuracy to which we can measure. We are not in a position to correct any fault or error until we know its nature and amount. In mass production, the measurements are made with some form of special and rapid measuring devices, many of them in the form of inspection gages. In production, corrections are made in the set-ups only when the product does not satisfy the requirements of these gages. In other words, the product follows the conditions permitted by the gages. Once the inspection procedure has been established, those conditions of the product acceptable to this inspection procedure supercede any strict geometrical or theoretical interpretations expressed or implied by the detail drawings. Under these conditions, one major purpose of the dimensional specifications on the detail drawings is to indicate or specify the inspection procedure,- including the design of the gages. At the same time, the conditions which these gages will accept must be those that will ensure an acceptable product.

The problem of dimensions is a complex one. In the first place, there are three kinds of dimensions that may be needed on a detail drawing. For one, there are constructional dimensions. These are needed to establish general forms and positions such as dimensions for patterns and castings, forgings, dies, irregular profiles, and the like, but these constructional dimensions are never measured directly in the course of continued production. These constructional dimensions never carry a tolerance.

Second, there is often need of some calculated dimensions for record or reference which are best preserved on the detail drawings. One example would be the pitch diameter of a gear,- a feature that does not actually exist on any single gear until it is meshed with its mate,- but which is needed to establish the tooth sizes and proportions. These calculated dimensions never carry a tolerance. Calculated dimensions should be given to as many decimal places as may be needed to obtain the required results. For example, the value of the taper per inch on a tapered feature may need to be expressed to five decimal places to ensure the needed results.

Third, most of the dimensions for machined surfaces are limiting dimensions, expressed as a size with a tolerance or as the extreme limits of size. These values should be expressed only to the number of decimals that will actually be used in their measurement. These limiting dimensions are the most troublesome ones, they are variables and not constants.

These limiting dimensions have several different kinds of conditions to control. The treatment that may be best for one kind of condition is not necessarily adequate to control a different condition. The nature of the variations may be quite different. Hence in establishing the limiting conditions for any feature of the component, careful consideration must be given to the specific kind of condition involved and the nature of the variations which they must control.

The limiting dimensions or conditions expressed on the drawings control the extreme sizes of whatever limit gages or other inspection facilities that are used. As noted before, the actual conditions of the product which is accepted by the limit gages must be such that a satisfactory product is obtained. It cannot be emphasized too strongly that the performance of the assembled product is the ultimate reference,- or Supreme Court,- which must pass upon and approve or reject every specific limiting dimension that is used.

At the start of new production, the detail drawings with their dimensions and tolerances represent the first best guess of the production design department as to what is needed to meet the operating conditions of the product. The task of dimensioning with tolerances is one that has no place for inexperience. The engineers engaged in this work must be experienced production men. They must know the available processes for the forming and shaping of materials, the order of accuracy and the characteristics of the surfaces which can be obtained and maintained and the relative costs of these operations. Zeal is necessary and desirable, but zeal alone is never an adequate substitute for ability and experience.

After production has started, more information about the whole production problem is available if we look for it. Hence the production design engineers must keep in constant touch with the progress of production, and make all necessary and desirable revisions and improvments to further the smooth and economical flow of satisfactory production. Production design is a continuing process,- a matter of evolution,- which is never finished as long as the particular product is being manufactured. To this end, it is a good practice to make the Production Design Department responsible for the finished parts inspection and all functional testing of the assembled product. Then they will never have the excuse of not knowing about the need for any specific revision.

To return to the problem of the limiting dimensions and their tolerances: these may have any one of four different conditions to control. These are as follows:

(a) Conditions of size
(b) Conditions of form
(c) Conditions of position
(d) Conditions of functioning, including assembly, production and performance.

Conditions of size are the most common and are the easiest to deal with. They consist of diameters, thicknesses, lengths and all other elementary surfaces which can be controlled by a single dimension. To meet this condition, a dimension with its tolerances or the two extreme limiting sizes are given. The largest part of the present practice in dimensioning with tolerances is concerned primarily with this condition of size. Attempts are made in too many cases to treat all other conditions in the same manner,- frequently with disasterous results.

Conditions of form should be self-explanatory. It may be the form of a cam, or of a gear-tooth, or of a machine handle which is a combination of radii and straight lines. The form can be defined with the necessary constructional dimensions. But any tolerances to control the permissible variations in the shape or contour of this form

cannot be expressed clearly by tolerances on these constructional dimensions. The allowable conditions of variation on one form may be quite different from those on another condition of form. One good and commonly used method of expressing the tolerances for such conditions consists of specifying a tolerance that specifies how much variation can be allowed in a direction normal to the form. This type of tolerance is sometimes called a "Metal tolerance". When plus, it allows more metal on the form; when minus, it allows less metal on the form. These conditions can be checked by a dial indicator gage that measures this variation by comparison with an accurate master profile which has been made in the gage shop from the constructional dimensions given on the detail drawing. For many small parts, it can be checked on an optical comparator against a master form which has been drawn to the proper enlargement on the screen of the comparator.

The specification and control of conditions of size and conditions of form cannot be combined if the most effective results are needed. For example, it may be a three inches diameter shaft for a high speed rotor that is to be manufactured. Its diameter may vary three or four thousandths of an inch without adversely affecting its performance in its bearing, but the form of its bearing surface must be cylindrical within a few tenths of a thousandth of an inch to obtain smooth rotation of the rotor. In such a case, limiting conditions must be given for the diameter (condition of size) and a separate tolerance must be given for the permissible variation in its form.

At this point it might be well to point out that no limiting specification of any kind should ever be expressed that will not be definitely measured in production. It is also significant to note that the most troublesome and perplexing problems of design and control that we have today persist because we have been unable to devise simple and adequate means to measure them rapidly and effectively. We must have some such means to measure them before we can express them as intelligible specifications. As a corollary, all dimensional specifications must be expressed in terms of their measurements. Despite this handicap of inadequate means of measurement, the parts must be manufactured, and here the limiting conditions are often expressed as the cumulative tolerance on some condition of size. The resulting gage is a functional gage, a condition that will be discussed later.

Conditions of position include the locations of holes, the spacing of gear-teeth and splines, conditions of symmetry, alignment, of conditions of eccentricity of several shoulders on a shaft, and so on. Here again we meet many different problems in different cases. Sometimes the local tooth-to-tooth variation is the critical one. In other cases it is the total or maximum cumulative error over a number of these spaces or steps that is the more critical one. Again, it may be a combination of both. In any event, the limiting conditions must be expressed, by dimensions or symbols or notes, in terms of the measurement that is to be used.

At the present time we do not have any simple and commonly-accepted method of expressing these conditions. In many cases we have resorted to some more or less lengthy notes printed on the drawings. The wording of these notes is not always as clear and definite as they should be. For the location of most of the holes, reliance is placed on the accuracy of the drill jigs and boring fixtures, and no positional

tolerance is expressed. This practice is adequate when the mating parts are made in the same plant and the different drill jigs and boring fixtures can be adjusted to each other so that the correct assembling and operational conditions are secured and maintained. But if mating parts are made in different plants, difficulties because of uncontrolled variations in conditions of position do sometimes occur.

In addition, in most cases a condition of position involves some definite base line or datum. Such datums must be surfaces. We cannot measure from a hypothetical centerline. But these datums are seldom mentioned in the notes that are used.

A simple positional tolerance symbol has been proposed,- one that also identifies all needed datums. This symbol method is adequate for a large majority of the positional conditions, but it has not been received with favor by many engineers. It is possible that they do not appreciate the importance of control for this condition of position. Their major expressed objection appears to be a reluctance to add any new symbols to the language of drawings. This reluctance I cannot understand. The language of drawings should always strive to reduce the methods of expressing dimensional information to the most concise form possible.

For example, the S.A.E. Automotive Drafting Standards give 26 symbols for various types and conditions of welding,- all new in comparatively recent years; and a new Surface Finish symbol with about 9 variations for different conditions; but the addition of two more positional symbols to help clear up the important positional tolerance problem appears to be the straw that breaks the camel's back. Here I think the philosophy of George B. Grant, expressed over sixty years ago, applies. He states that the old -- "holds its position primarily because of human inertia, or the natural reluctance of the average human mind to accept a change,- particularly a change for the better".

Conditions of functioning include all conditions that do not fall into any one of the three categories. It is much the same as the item: "Incidentals (please explain)" that appears on many very detailed expense-account forms. Many times some condition that involves a complex inter-relationship between several features must be controlled. The need for such control often becomes evident only after the initial production of a new product is well under way. Some simple test or functional gage is devised that gives the needed control, but there appears to be no way to express it simply and effectively in terms of dimensions and tolerances. In such cases the functional test or gage itself becomes an integral part of the dimensional specifications for the particular part.

As one common example of a functional gage we will consider the gages used for screw threads. This type of feature involves the diameter of the screw thread (condition of size), the form of the thread (condition of form), the squareness or alignment of the thread form (condition of position) and the pitch or lead of the threads (condition of position). All three of these conditions are intimately related to each other. Each condition has its own particular and sometimes critical functional purpose. In such a case, the most effective control would be obtained by measuring each condition independently. But thus far no one has devised any simple, rapid and effective means to do this. The best solution that we have found to

date is to give a tolerance on the pitch diameter (a condition of size) and specify that this tolerance is a cumulative one and includes all errors of form, or thread angle, and lead as well as diameter. The resulting "go" thread gage is a functional gage that ensures primarily that the threaded parts will assemble together. For critical conditions the tolerances are made smaller,- not because the condition of size itself is critical or important but because with the smaller diameter tolerance, the lead and thread form must be made more accurately in order to meet the conditions defined by the gages. There is much room for improvement here.

Functional gages are an old story in the manufacture of guns, large and small. A dummy breech mechanism is always provided for the manufacture of any field gun. In effect, this is a prototype model made in the tool room to the detail drawing specifications. It is a functional gage into which all parts from current production must assemble and operate correctly.

A few manufacturing plants make one or more manufacturing models of a new product in the tool room or model shop to prove the adequacy of the production design. These serve several purposes. One model may be saved to use as a functional gage to prove the correctness of the first tool-made parts at the start of new production. This practice often catches many troublesome errors before production is too far along. The other models are often used for endurance test running and overload tests thus revealing probable operating troubles earlier in the course of production than they would otherwise become known, and giving the manufacturer an opportunity to correct them before they reach the hands of their customers.

All this leads to another important problem and this is the formulation of some system or language of drawings that will be generally acceptable to the production industry. I have my own ideas or notions, subject to change as I learn better,- and other people have theirs, much of which is probably better than mine. Some adequate system is urgently needed now.

EXPERIMENTS IN THE CHEMICAL PLANT

James A. Mitchell
Tennessee Eastman Company

In the operation of a chemical process there are many times when it is necessary to run a production scale experiment. Chemical process advances learned in the laboratory, or semi-plant, must be tested for final approval in actual production. There are also cases when certain conditions cannot be simulated in the laboratory and experiments on these factors must be run in production. Sometimes, as we shall find, it is most economical to do certain types of experiments right in production.

Planning, conducting, and interpreting these experiments is a task generally falling on one or more of the chemical engineers or chemists in the production, development, or quality control staffs. I have always found the men experienced in this work definitely impressed with the difficulty of planning and interpreting production experiments so that clear, unambiguous answers are obtained with a minimum of cost in time and money and with a minimum of risk to the production involved. One phase of the skill required in planning good experiments is, of course, the knowledge of the chemistry, engineering, and practical conditions which are a part of each individual chemical process. A second phase is the science of experiment planning and interpretation of data which is common to all experimentation. It is this second phase that I plan to discuss with special reference to plant scale experiments.

Plant scale experiments are only a part of the chemical development program and must be coordinated with semi-plant experimentation and laboratory experimentation programs. It is part of the problem of the experimenter to decide on what scale experiments should be tried and what sequence of experiments can most efficiently and quickly solve the problem. Sometimes we will plan small production scale experiments to be followed by larger ones. An experiment's real value lies in its relation to advancing the experimental program to its conclusion. (See Figure 1).

In general, no experiments should be conducted on a plant scale unless sufficient knowledge of the experimental conditions is in hand to be assured that no serious hazard will be placed on the product or the production process. Data on these points are usually obtained from previous experience or semi-plant or laboratory tests.

First, a brief review of the science of experimentation is in order. An experiment has been defined as "a trial conducted to verify or deny a hypothesis". It is an attempt to make the process "talk", or at least answer some questions. We must remember, however, that we can only ask certain kinds of questions and can only receive certain limited answers. We cannot design experiments which will answer questions like, "What catalyst will give the best yields?" or "What is the cause of haze in this plastic?" We can only ask, "Will catalyst A give a better yield than catalyst B?" or "Will increasing the temperature 10°F. at point X affect the haze?" In order to test the answers to these questions

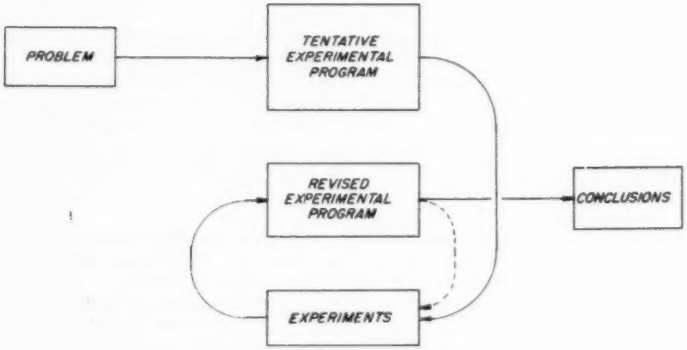


FIG. 1

by rigorous statistical mathematics, it is customary to put the question to be answered in the form of the so-called "null hypothesis". This is simply a statement of the hypothesis on the side of the question assuming no difference in the effect of the two levels of the factor. For example:

(1) There is no difference in yields using catalyst A or B.

(2) Increasing the temperature at point B by 10°F. will not change the haze in the plastic.

The answers which the experiment can give us are therefore of the following type:

I-A. Insufficient evidence to cause us to reject the null hypothesis. In a simple case, this would mean that the difference in results of the two levels of the factor was less than the error of the experiment.

I-B. Sufficient evidence to cause us to reject the null hypothesis. Again, in a simple case, this evidence is a difference in the results of the two levels of the factor which is sufficiently larger than the error of the experiment that the probability of observing such a difference from chance variations is below some critical value.

II. A measure of the error of the experiment.

Now, if we get answer I-A, it may be that there actually is no difference in the effect of the two factors; or it may simply be that the experiment is not precise enough to detect it. We can calculate from the error of the experiment the maximum difference which could exist at a selected probability and then decide if this difference is of any practical importance. Note that we never "prove" the null hypothesis. We can only state that there is insufficient evidence to disprove it.

If we get answer I-B, we can also determine the precision of the experiment and then by assuming a certain probability of risk we can determine within what limits the actual difference can be depended to lie.

This, then, is how an experiment can answer our questions. A good experiment can tell us the most with the least time and cost and the least risk to the production process. There are many possible plans for conducting plant scale experiments. Traditionally we think first of the method--vary one factor while holding all others constant. It is obvious that is not practical under plant conditions and indeed under any conditions, so the first method is really--vary one factor while the other factors follow existing variation.

This general type of experimental method has had a lot of use and in the hands of experienced men has produced results. However, as generally used this method has led to a high percentage of erroneous conclusions primarily, I believe, because of two shortcomings.

(1) No measure of the error of the experiment is usually obtained.

(2) It provides no assurance that the difference observed is not due to other factors.

It is possible to design experiments based on this general method which do use modern statistical methods and overcome some of the objections. This technique we call the control chart experiment. The philosophy of this method is to study the amount and consistency of the variation of the characteristic in a process, then to change the factor in question, observing the change in level and variability of the characteristic. Consistency of the variability must be established before the results of the experiment can be relied upon. (See Figure 2).

Another experimental plan is the randomized factorial experiment. In this plan several factors are simultaneously varied at a number of levels, the experiments being run in an order determined by chance. The effects of each of the factors are separated by a technique called variance analysis. This method and its modifications are the most powerful and efficient of the experiment techniques, but for two reasons I am going to confine my discussion at this time to the control chart experiment. First, I want to spend most of the time discussing some of the more fundamental angles of the problem, and there will not be time to treat the factorial experiment method. Second, experiments in the production process fall easily into the control chart experiment method and engineers should become familiar with this technique before tackling factorial experiments.

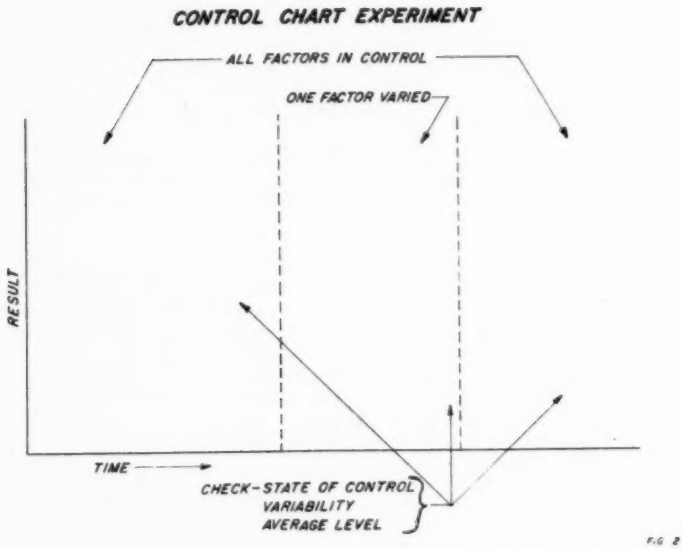


FIG 2

Let us look at an example. In the Cellulose Esters Division of the Tennessee Eastman Company we make many types of cellulose esters for the film, plastics, yarn, and coatings industries. The chemical and physical properties of each of our cellulose ester types must be closely controlled within narrow limits and an intensive program of control and development is continually in progress. Plant experiments have their place in the program, and the trial of a new source of cellulose in the manufacture of cellulose acetate butyrate for plastics is a typical example.

The general purpose of the experiment should be clearly set forth. In this case it was "To determine what effect the substitution of cellulose "A" for regular cellulose would have on the quality of cellulose esters used for plastics".

A list of all product characteristics which are part of the general requirement should be made. From this list should be selected those characteristics of the product which are to be checked critically and those on which simple inspection will suffice. It is especially important to have good quantitative measurements on the critical quality characteristics. Sometimes you encounter an important quality characteristic which is difficult to measure quantitatively. It is extremely difficult to experiment under these conditions, and it is worth a great deal of effort to find a method of measuring the characteristic before experimenting. In this case we will select for further study one of the critical characteristics of the product--plastic color. This is a difficult quality to measure as it really amounts to a practical use test. The cellulose ester is compounded with a standard plasticizer and formed into plastic, then subjected to a standard heat treatment greater than

would be encountered in normal molding practice. The color of the plastic piece is then measured against standards on an arbitrary scale.

It is next necessary to make some quantitative decisions defining the general purpose of the experiment; for example, What change in the plastic color would be a minimum change of practical significance? It is also necessary to decide at what level of probability we will reject the null hypothesis. In this case, it was felt that any change in the average heat test color by one unit on our arbitrary scale would be a change of practical importance. It was also decided that a 1% risk of error in the conclusions would be the most that could be accepted.

The past record of the quality characteristic under investigation should be examined and a control chart set up on this data. This will provide information of the average of the process, the variation of the process (the standard deviation), and the state of control (whether the data remains within control limits or not). If the data tends to be in control, then we have reason to believe that the variability due to all of the factors in production is controlled within limits and a valid comparison can be made with an experimental condition of one factor changed. If the data tends to be out of control, then there is a considerable risk in judging the effect of experimental conditions. This data also provides a measure of the variability of the process and allows us to estimate the size of the experiment which will be required to fulfill the conditions set forth. In this case, we find from past records of several hundred results that the process is in control and the variability of the results have a standard deviation of 2.0. With this amount of data, the observed average and standard deviation can be considered as standard values essentially without error.

The quantitative requirements of the experiment are therefore:

INFORMATION FOR DESIGN OF EXPERIMENT

Null hypothesis - That the plastic color of the cellulose ester made from cellulose "A" is the same as that made from regular cellulose.

Probability level of significance for rejecting the null hypothesis 1%

Minimum difference of practical significance . . . . . 1 unit

Process Average 34.5

Process Standard Deviation 2.0

State of Control . . . . . .In control

The "t" test can be used to estimate the number of experimental batches that will be required, assuming the same variability in the experimental period. If we have over 100 results in the past record, the formula in a reasonable approximation is $n = \left(\frac{\sigma t_\infty}{D}\right)^2$

where σ = the standard deviation.

t_∞ = t value for infinite degrees of freedom at the selected probability.

D = the difference in average to be detected.

$$n = \left(\frac{2.0 \times 2.58}{1.0}\right)^2 = 26.5$$

With this information, it is possible to design the experiment so that it will be no larger than necessary. It is also possible to see in advance if it is impossible to fulfill the requirements of the experiment with the practical size of the experiment allowed by production limitations. In this case 35 batches were selected as a reasonable practical number.

Conduct the experiment and examine the data immediately before, during, and, if possible, after the experiment by control chart methods. Separate control limits should be determined for the experimental period if points fall outside of the regular process limits. The data for this example experiment is shown in Figure 3. Note that there has been a

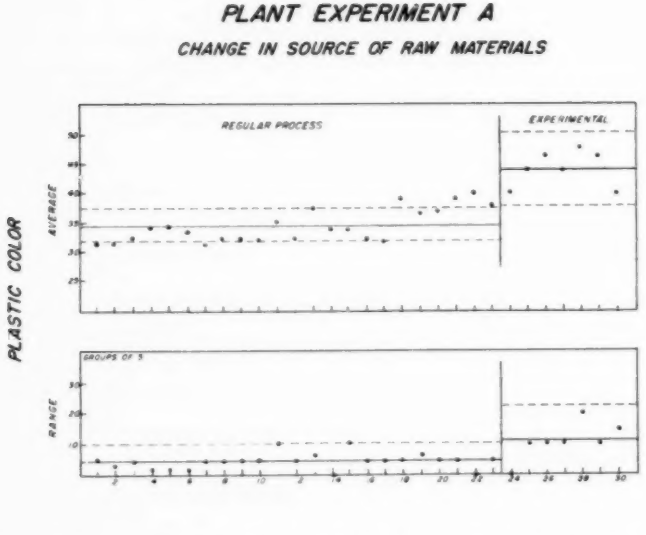


FIG 3

definite increase in the average plastic color. The comparison of data before and during the experiment would be as follows:

	Regular Production	Experimental Run
No. of groups	100	7
n in a group	5	5
n total	500	35
average	34.7	44.0
average range	4.7	10.7
standard deviation	2.0	4.6
state of control	lack of control just before experiment.	in control

The fact that the data immediately before the experiment shows lack of control throws considerable doubt upon the interpretation of the experiment. You will note that there appears to be a rising trend in the plastic color just prior to the experiment. The points out of control on the control chart show that this trend is greater than can be due to chance alone. Although the experimental data is at a higher level than the previous production, this information is questionable since the rising trend prior to production may simply have continued into the experimental period. We therefore must conclude that although the most probable conclusion is that substituting material "A" definitely increases the plastic color and the plastic color variability, there is doubt about the other factors remaining within control, and further experimentation or investigation is necessary to really settle the question. Subsequent study of production conditions did provide a possible reason for the changing level of plastic color. With later experiments it was shown that substituting material "A" for regular cellulose did increase the plastic color, but only about half as much as this data indicated.

Another example of this same type of problem is the data shown in the following example. In this case we are also measuring plastic color, but a different type of cellulose ester is involved. On the color scale used for this plastic a difference of 10 is felt to be the just practically significant point and we are still working with a 1% risk of error. The quantitative requirements of this experiment are therefore:

Null hypothesis - That the plastic color of the cellulose ester made with treatment "Z" in the process is the same as the regular process.

Probability level of significance for rejecting the null hypothesis 1%

Minimum difference of practical significance . . . 10 units

Process Average 180

Process Standard Deviation 20

State of Control . . . . . In control

The required number of results to determine the minimum difference of practical significance at the probability level given is

$$n = \left(\frac{\sigma\, t_{\infty}}{D}\right)^2 = \left(\frac{20 \times 2.6}{10}\right)^2 = 27$$

The data just before and during the experiment are shown in Figure 4. Here the control chart data shows the process to be in control before and after the experiment and also during the experiment with different control limits. In this case we feel it is safe to proceed with some experimental conclusions. The data for this experiment may be summarized in the following table:

	Regular Production	Experimental Run
No. of groups	100	16
No. in groups	3	3
total n	300	48

	Regular Production	Experimental Run
average	182.7	156.8
average range	36.0	24.7
standard deviation	21.3	14.6
state of control	in control	in control

It is generally regarded that about 25 points in control show that a process is in a state of control with a reasonable degree of assurance. In this case there are 100 points in the regular process and 16 for the experiment. Since the process was within control for some period of time before this data, this provides reasonable assurance of control in the process.

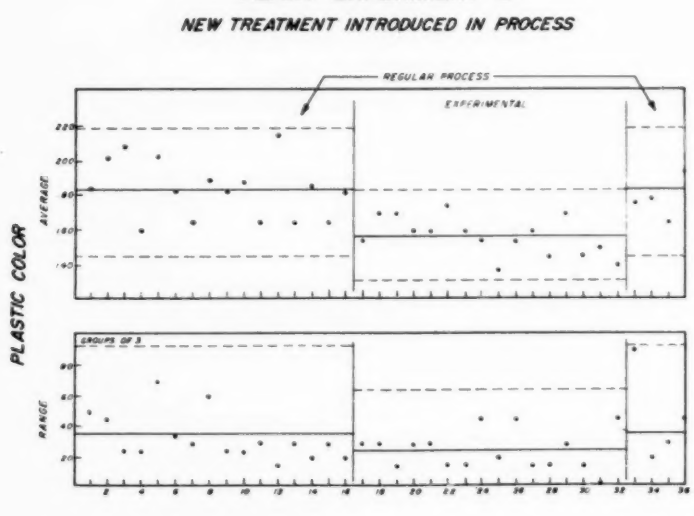


FIG 4

First the difference observed in the variability of the two processes should be tested by the F test to determine if it is significant. In this test the ratio of the variances is compared with an F value obtained from statistical tables. If the calculated F ratio exceeds the critical value given by the table for our selected probability the variances are shown to be different.

$$\frac{\sigma_1^2}{\sigma_2^2} = F \qquad \frac{(21.3)^2}{(14.6)^2} = 2.13 \quad F(n_1 = 300, n_2 = 48) = 1.9$$

(From table P = .005
Used as P = .01)

The variation of the regular process is therefore significantly greater than the experimental process.

Now that we have compared the two variances, we are ready to compare the significance of the difference of the two process averages.

If the two variances had not been significantly different, it would be reasonable to assume that the variance of the process had not changed. The regular process variance would then be used in the following approximate formula:

$$t_\infty = \frac{X' - \bar{X}_2}{\sigma'/\sqrt{n}}$$

Where X' is the regular process average, $\bar{X}_2$ is the experimental average; σ' is the standard deviation of the regular process; t_∞ is the t value for infinite no. of degrees of freedom; and n is the number of values in the experimental average.

When the two variances differ significantly the following formula is used:

$$t_n = \frac{X' - \bar{X}_2}{\sigma_{x_2}/\sqrt{n}}$$

Where the notation is the same as above and σ_{x_2} is the standard deviation of the experimental data, and t_n is the t value for number of degrees of freedom in the experimental average.

In this case the second formula is used.

$$t = \frac{182.7 - 156.8}{14.6/\sqrt{48}} = \frac{25.9 \times \sqrt{48}}{14.6} = 12.3$$

$$t_{47}(P = .01) = 2.7$$

So the calculated t value is greater than the critical t value from the table and the improvement in quality is significant.

The indicated improvement is from 182.7 to 156.8, but the second figure is not highly precise. Confidence limits around this estimated average which will include the true average may be calculated if we will decide on a probability of error. Deciding that we want to predict these limits correctly 99 times out of 100, we select a P = .01.

The limits of the mean are therefore:

$$\bar{X} \pm \frac{t\sigma}{\sqrt{n}}$$

$$156.8 \pm \frac{2.68(14.6)}{48} = 156.8 \pm 5.6$$

The conclusions of this experiment may be set forth as follows:

Null hypothesis that the variance of the experimental process is the same as the production process is rejected.

Null hypothesis that the average of the experimental process is the same as the production process is rejected.

The best estimate of the effect of the experimental process is

Plastic Color

	Production	Experimental
Average	182.7	156.8 ± 5.6
Standard Deviation	21.3	14.6

In the examples given sufficient data was available from the production process to consider the average and standard deviation of the production process as being essentially without error. When this is not feasible, the formulas used must be modified as described in many statistical texts.

There are a number of improvements in this technique that will occur to those experienced in modern experimental methods, that plant engineers should know. First, the experimental process could be randomized right along with production checks. Such technique provides additional assurance that differences observed are not due to other factors. Secondly, when the other factors cause out-of-control data regularly, the experiment may be broken down into randomized blocks corresponding to more uniform conditions occurring for shorter periods of time, for lots of raw material, for different pieces of equipment or operators. This technique can greatly reduce the experimental error and improve the discerning power of the experiment. In some processes, however, it is difficult to arrange for the frequent change in conditions necessary for a randomized experiment and in these cases the simpler control chart experiment can be valuable to even the experienced experimenter. Even in these cases, it will sometimes be found advisable to randomize the testing of the product if the testing procedure is an important source of error. For example, if the production material produced in a control chart experiment must be further treated in the pilot plant for evaluation, complete randomization and block grouping of the pilot plant tests is very much in order. Whenever part of an experiment is not randomized it is essential to know that the process involved is in a state of control in the statistical sense. This is equivalent to testing the data for independence in replication.

In some production processes it is impossible to have no greater variation over a long period of time as compared to shorter periods. Data from such a process shows up as lack of control on the control chart and the control chart method of experimentation is not directly applicable to this data. The randomized block experiment is most efficient in these cases.

This, then, is the control chart experimental method. It fits well into the regular production process and improves the success of simple production experiments by providing a measure of error of the experiment so that chance fluctuations in the results will not be interpreted as experimental results for a high percentage of the experiments. It also provides for a method of assuring the experimenter that the variations observed are not due to some other process variable. If followed with intelligence, the experimenter will have a high percentage of successful experiments.

APPLICATION OF CONSOLIDATED COMPUTER TO THE SOLUTION OF A METALLURGICAL CORRELATION PROBLEM

Donald S. Leckie and Richard F. Cain
Republic Steel Corporation

OBJECTIVES OF PAPER

The use of the methods of correlation analysis in the investigation of industrial problems has become of increasing interest in recent years. The methods themselves are not particularly new; many sciences have used them in one way or another for a long time. Industrial "trouble-shooters", however, are now adapting the more advanced analytical techniques to supplement their established methods of data analysis.

One might ask, "What information does the application of multiple correlation (or multiple regression as others call it) to a set of data furnish to the investigator?" Answered briefly, the method develops an equation with which some quantity, or variable, can be estimated, and a measure of the closeness of the estimates to the actual values can be computed. The method also yields a set of factors which indicates the expected change in the dependent variable with a given change in each of the included independent variables. The reliability of conclusions may be estimated to the extent that the data are representative of the usual conditions affecting the process. If the data represent a special set of conditions any conclusions will be valid for that set.

It may properly be suspected that the development of these equations is a long, tedious process if several variables are to be included with enough data to be representative of the many conditions affecting a process. The purpose of this paper is to show how we have been able to reduce many of the tedious mathematical computations to a routine, in order that the time of the investigator may be more profitably spent in planning the investigation, in interpreting the results, and in checking predictions and estimates, rather than chasing computational mistakes and sweating out answers.

Specifically, we shall show how a combination of IBM computing equipment and the Consolidated Computer has been used to transform the handling of data to routine machine operations. To illustrate this, a problem has been chosen which permitted predictions in the interpretation phase, and furnished subsequent checks on the accuracy of the predictions. This problem involves the study of the production rate of a single furnace in one of our Open Hearth shops. The effects of variations of a number of operating variables upon Tons per Hour as a measure of open hearth performance were being sought.

This specific study is not considered to be an ideal textbook model of a correlation study. It was thought that it would be more useful to present a study where compromises with the ideally desirable have been made to illustrate the problems and decisions which face an investigator when he undertakes such a study. An example is the use of Tons per Hour as a dependent variable. From a mathematical standpoint, tons per hour is exactly determined when tons produced and heat time are known. A more direct approach, therefore, would be to study the effect of operating variables on these two dependent variables. Since tons per hour is a ratio of two variables in this specific case, we shall be fitting a linear

model to a non-linear variable. Tons per hour, however, is a traditional measure which open hearth operating people are psychologically prepared to discuss. Also, it has the effect of combining into one variable the two quantities, heat size and heat time.

When we include scrap charging time as an independent variable we cannot help but have a resulting correlation, because heat time is the sum of this specific time period plus others. We desire, however, an estimate of the relationship between the dependent variable and other independent variables with the effect of variation of scrap charging time excluded, and we achieve this effect by including scrap charging time as an independent variable.

In this presentation, therefore, we shall claim neither that this is the ultimate in open hearth studies nor that we have developed a universal method to be applied in all cases. Nor do we present the conclusions of this study as the optimum in open hearth operating practice. We can, however, show that our methods in this case enabled us to make predictions which were reasonably borne out in subsequent practice. In this case, at least, our compromises seem justified by results.

PLANNING THE STUDY

Several questions arising from discussions of previous studies and shop practices suggested a study, by multiple correlation, of a rather large number of operating variables to determine what variations in furnace practice are most responsible for large variation in tons per hour from heat to heat. The development of effective controls of the more influential variables with consequent improvement in furnace performance was envisioned.

Considerable careful planning with open hearth supervisory and technical personnel preceded the gathering of data. Special report forms were devised to facilitate the reporting of data not included in routine heat records. Operating personnel, such as melters, furnace helpers, stockers, and recorders were briefed on the aims of the study and the importance of accurate data. It was decided to study the performance of one 190 ton gas and oil-fired open hearth furnace for approximately the extent of a furnace campaign. This furnace normally charged scrap and about 40% hot metal and produced various grades of carbon and alloy steels.

GATHERING THE DATA

A list was compiled of operating variables which included those considered to be important and others thought to have possibilities of importance. The data applying to such variables were reported and tabulated by open hearth personnel under the supervision of the Quality Control Department. In the interest of some economy of effort, 23 variables (Table No. 1 in the Appendix) were selected for correlation. These were discussed with the operating people who agreed that the most important and most promising measurable variables were included. Under other conditions and with other aims different variables might be used. Exceptional heats (roof patches, forced tap, etc.) were omitted leaving 125 heats produced over a period of about three months. Since no separation was made for grade, the data represent the average range of grades generally made in the furnace over a period of time.

PREPARATION OF DATA FOR COMPUTER

After a set of data has been accumulated, an investigation enters the cumbersome computational stage. To keep such an investigation from dying of its own inertia, these calculations must be done quickly and accurately for top efficiency. Accuracy must be stressed especially. The most disheartening task for the investigator is to seek mistakes after inconsistencies have become apparent in the computations. Also, unless there is absolute assurance of accuracy, the nagging suspicion will always be present that a mistake may have been made which did not result in any inconsistency observed during the process of calculation.

Since it is beyond the scope of this paper to present the details of a multiple correlation problem, from this point on reasonable familiarity with such procedure must be assumed.

The first step is to code the data to a suitable number of digits so it can be handled easily in the calculating machines. Our practice is to code to a maximum of three digits, although often two digits will prove to be satisfactory. Coding may be done by adding or subtracting a constant, or by multiplying or dividing by a constant, or both. The smaller the coded numbers are, the smaller the resulting squares and cross-products will be, thus providing more convenient numbers for future computations. Care must be taken, however, not to code data so coarsely that significant variation is lost. For instance, tons per hour could be rounded to the nearest unit. This would, however, eliminate the tenths which are certainly significant, since tons and heat time are routinely reported to three significant figures. Consequently, tons per hour were coded (Table No. 5) by multiplying the column of data by ten to get whole numbers, then subtracting a constant, seventy, to reduce the size of the coded numbers. All coding must be accurately done or the investigator will be analyzing a set of data which is different from the one which was so carefully collected.

The coded data for this particular study were tabulated in twenty-three columns on ruled sheets of paper, and then these sheets were given to the Tabulating Department where the operators key-punched the data into cards on an IBM No. 024 Numerical Key Punch. The cards were set up to contain twenty-three 3-digit fields, or 69 digits. This master deck of cards was verified for correct punching and then run through the IBM No. 402 Tabulator to sum the columns and record the sums. These sums were checked against a set of column totals obtained on an adding machine in the Quality Control Department as a final check on the key-punching.

The Tabulating Department has a method of computing the sums of squares and cross-products of columns by duplicating the master deck on an IBM No. 514 Reproducing Punch Machine and then running the decks through either a No. 602 or a No. 604 Multiplier. The multiplier accumulates the sums of the squares and cross products as a deck runs through the machine, and the accumulated answer is punched in a summary card following the deck. The multiplication is done in such a way that each sum of squares and cross-products is calculated in two independent ways for checking. The summary cards are then tabulated on an IBM No. 402 Tabulator for a listing of the sums of squares and cross-products as well as the set of checks. A summary of the sums, squares, and cross-products of eleven important variables is included in Table No. 2 in the Appendix.

At this point some people immediately set up and proceed to solve the normal equations. Others, however, compute simple correlation coefficients (r's) from the preceding data and use these r's in the normal equations. There are certain mechanical objections to the direct use of squares and cross-products. An example is the number of significant figures that must be carried to preserve the required accuracy in the solution. Also, the computation of auxiliary information such as R^2 (the multiple coefficient of determination) is simpler with the data transformed, so we have always considered the extra operation worth while.

For a large set of data such as this the calculation of the r's is a routine operation adaptable to IBM equipment. Table No. 3 in the Appendix displays the formulas and the tabular form used. The Quality Control Department fills in Columns 1, 2, 3, 4, 5, 7, and 8 on the form. Cards are punched from this data, and the IBM Department computes the numerator and the denominator product of the formula in Columns 6 and 9. The division of Column 6 by Column 9 may be performed by IBM equipment or on a desk calculator.

The calculation of r's is done by groups, with the first group containing the correlation of a selected variable with all the others, the second group containing the correlations of a second variable with the other variables less the first selected variable, etc. The computation may be checked by summing the columns and performing the calculations as indicated in Table No. 3. The result of applying these calculations to our data is shown in Table No. 4 which lists the r's to be used for the solution of the normal equations.

In some textbooks the emphasis at this point is on the direct solution for the coefficients by some method such as Doolittle's or Dwyer's. For explaining the operation of the computer, however, the simultaneous solution of equations will have to be stressed. These equations are of the form:

$$\beta_1 + r_{12}\beta_2 + r_{13}\beta_3 + \cdots + r_{1N}\beta_N = r_{01}$$

$$r_{12}\beta_1 + \beta_2 + r_{23}\beta_3 + \cdots + r_{2N}\beta_N = r_{02}$$

$$\vdots$$

$$r_{1N}\beta_1 + r_{2N}\beta_2 + r_{3N}\beta_3 + \cdots + \beta_N = r_{0N}$$

The r's are known and the Betas are the unknowns whose solutions are required. For "n" independent variables there are "n" unknown Betas and "n" equations. Any one who has tried to solve four or more simultaneous equations algebraically has some idea of the extent of the tedious calculations. The Consolidated Computer solves such equations by the entirely different method of successive approximations. This machine is the Model 30-103 Computer manufactured by the Consolidated Engineering Corporation, Pasadena, California.

DESCRIPTION OF COMPUTER

A diagrammatic picture of the front panel is shown in Fig. A. This computer is an analog, rather than a digital, type. The computer houses a large turret (A) which contains twelve rows of 13 helipots (C) (potentiometers) each, and which is revolved to the proper row at the desired time by the hand wheel. (B). The coefficients of each equation (r value) are set on the helipots in an equivalent row on the turret. Thus, the machine has a capacity of twelve linear equations with twelve unknowns. A system of pilot lights (K) indicates which circuit components are connected for various settings of the operating switches.

The "r" coefficients are set on the computer by means of the decade switch (D) which has four dials, each numbered from zero to nine. Any four digit number may be set on the dials and there is a special provision for setting the number 1.0000. Since we have always followed the practice of computing the correlation to four decimal places the machine has the proper capacity for our use. We have found in the past that r's computed to more than four decimal places do not seem to increase desirably the accuracy of the solution in view of the possible inaccuracies in the basic data. If computed to less than four places the resultant accuracy may sometimes be less than desired.

Any selected "r" is entered in the machine by revolving the turret to the proper row and by setting the column selector switch (E) to the proper column. The number is set on the decade switch and then the proper helipot, indicated by pilot lights, is turned in the direction indicated by the neon glow-lamps (G) until a circular pattern on the cathode ray tube (H) closes to a horizontal line. The r's for the dependent variable are always set in the last column of helipots on the right-hand side of the turret.

After the entire set of equation coefficients has been set up on the turret it is possible to check them easily. The computer is now ready to solve for the Betas. These Beta values are to be found on the bottom row of helipots (F).

The solution is achieved by a method of successive approximations, the mathematical explanation for which is considered to be inappropriate here. The computer is constructed so that the products of the r's on the turret and their respective Betas are summed and compared with the dependent r's which are on the right of the equal sign in the equations. The Betas are adjusted to balance the equation as indicated by closure of the circle on the cathode ray tube. Repeated approximations are made until a set of Betas is obtained which balances all equations. This set is the required solution of the equations. The solution is obtained in orderly fashion and successive approximations result in succeedingly smaller corrections to the Betas.

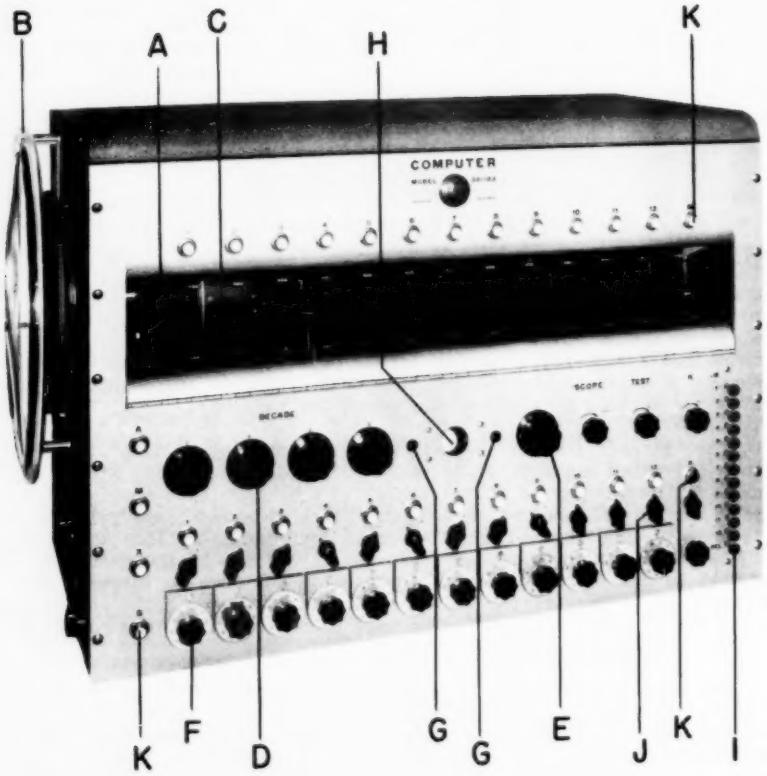


FIGURE A

A - Turret
B - Hand Wheel
C - "r" Coefficient Helipots
D - Decade Switch - 4 digits
E - Column Selector Switch
F - Beta (Unknown) Helipots
G. - Neon Glow Lamps
H - Cathode Ray Tube
I - Polarity Push Buttons
J - Polarity Switches
K - Pilot Lights

The computer is equipped with a set of polarity push buttons (I) and a set of polarity switches (J). These are both necessary to set the correct signs of the "r" coefficients for each equation as it enters the computer circuit during the solution cycle. The values of the Betas are found by balancing the decade switch against the Beta helipots, which may be selected in turn by the column selector switch. The Betas can be given an auxiliary check by substitution in any or all of the original equations on a desk calculating machine.

The literature (#1, Bibliography) with reference to the computer gives instructions for arrangement of the data for optimum speed of convergence to the solution. Under ordinary operating conditions, the coefficients for a twelve independent variable problem can be set up and checked in approximately 45 minutes and the solution may take from 10 to 45 minutes, depending on speed of convergence. This is quite a saving in time over the usual computational methods. An occasional problem will require an hour or more for solution, but such problems generally contain some statistically or mathematically undesirable condition, such as high intercorrelation among the independent variables.

One of the most valuable features of the computer is that once the r's are set up on the computer, different combinations of the independent variables may be studied by omitting the proper columns and equations in the solution. Also, a different dependent variable may be conveniently substituted on the turret to solve any combination of the independent variables with a new dependent variable. Regarding the set of correlations of the independent variables as a square matrix, the inverse of this matrix can be computed and the errors of the Betas computed from the diagonal terms. These are a few representative examples of the versatility and the value of the machine.

CALCULATION WITH THE COMPUTER

We have said that the machine has a capacity of twelve independent variables, and we have indicated that we had 23 variables in our set of data. This certainly exceeded the capacity of the computer. There are methods for extending the capacity of the computer to accommodate more than twelve variables. In this case, however, we did not try to solve with all 21 independent variables simultaneously, but used the following method:

The 21 independent variables were divided into two groups, one of eleven commonly studied variables whose effects had been studied before, and the other ten variables which had not been studied before because of a lack of adequate data, but in whose effects we were especially interested.

The first group was correlated with tons per hour as the dependent variable and the resultant coefficient of multiple determination, R^2 was .7087. Using a test developed by R. A. Fisher (#2, Bibliography) the three weakest variables with respect to contribution to R^2 were eliminated. The new solution with only eight independent variables yielded an $R = .7058$, only slightly lower than before. We then concluded that in this group the eight selected were the most important, and reserved them for further study.

When the second group was tested in a similar manner, all ten independent variables yielded a low R^2= .1624, and removing five of the ten variables resulted in an R^2 = .1508. The remaining five, together with the eight selected from Group 1, gave a total of 13 variables and these were studied by using the eight from Group 1 and varying combinations of four of the five from Group 2. It was finally determined that only two of the five from Group 2 were significant and the resulting R^2 with ten independent variables was .7217.

Thus, 10 of the 21 independent variables indicated that they were important enough in their effect upon tons per hour to merit further investigation and attempts toward increased control. It should be stressed that no conclusion can be made that the discarded 13 are totally unimportant in their effects upon open hearth production. The seeming lack of relationship may be due to limited ranges of variation in the practices during the data period, errors arising from sampling of tests, inadequate measurements, etc. Some of these relationships could be found, perhaps, with an experiment planned specifically to measure a particular variable effect.

Table No. 5 summarizes the pertinent results of this study and lists the simple correlations of the several independent variables with the dependent variable (r's), the Beta values, the t-Tests of Significance of the Betas, the R^2, and the Standard Error of Estimate (S_{x_o}) of the dependent variable. For comparison, the Betas calculated by the Computer and by the abbreviated Doolittle method are shown. Note that the agreement is close between pairs of Betas computed by the two methods. The Betas from the computer solution are of sufficient precision to be used for estimating purposes, with results well within the possible errors in the original data. The t-Test given is the ordinary statistical test of significance and is used to determine if the coefficient is so different from zero that it is unlikely to be that large due to random causes alone. The 1% and 5% significance levels are given and may be compared in the usual way. A majority of the variables will be seen to be highly significant. The estimating equation is also shown. This completes the laborious computation section of a study and, if the calculations are free of mistakes, the equation yields the least-squares estimates of the dependent variable with a predicted error or variance.

INTERPRETATION OF RESULTS

The results were now ready for the very critical phase of interpretation and presentation to operating personnel. In many industrial applications, doubts of the multiple correlation technique have arisen due to insufficient care in interpretation. In our organization, before interpretations are released to operating supervision, a review panel carefully evaluates the study both statistically and also from the standpoint of practical experience and prior knowledge. Conclusions which may not make sense at first sight are studied to determine if they might be possible under certain conditions, if they might have appeared due to an omitted factor, or if they might be due to incorrect data, etc. It should be stressed that we never immediately discard this type of statistical result merely because it fails to make sense or conform to previous opinion. It will always be possible to find new relationships in most processes, and many times it is desirable to collect more data and experiment further to check such an "absurd" conclusion for repetition.

It should always be recognized that the relationships fitted to the data by this technique have been linear, although in many cases the relationships are known to be curvilinear. The linear estimates, however, are a first approximation, and frequently the curvature is so slight in the range of the variable studied that the linear approximation is a very good one. Methods are available in the literature for investigation of curvilinearity after the problem has reached this stage. Also, it is important to remember that simple correlations and even multiple correlation results do not necessarily yield cause and effect relationships. Sometimes these relationships may be symptomatic as, for example, the significant relationship of the variable Burned Lime Added with Tons per Hour, which reflects a condition of the slag-bath chemistry. Any "treatment" must apply to the condition of the slag-bath rather than to the symptom (Burned Lime Added). In any event, cause and effect is demonstrated only when the results of actual process changes lie within the region of predicted results.

After the results of the correlation were digested by the reviewing panel, a meeting was held with open hearth operating supervision which included all foremen. Instead of presenting the mathematics and formulas, the results were presented by listing the probable effects on tons per hour for an attainable change in the average of important variables. In total, the changes recommended for the variables indicated a probable increase of .99 tons per hour from an average of 11.92 to 12.91. Some typical conclusions were:

(a) Increasing average heat size five tons per heat would mean a probable gain of .250 tons per hour.

(b) A reduction of one-half hour in scrap charging time would mean a probable increase of .153 tons per hour.

(c) A decrease of 20 minutes of non-productive time (bank and bottom-making delays) would mean a probable increase of .166 tons per hour.

(d) The study had shown that an average of 4.2 boxes of burned lime was added during the heat and if this amount were reduced by 2 boxes, on the average, the probable gain would be .241 tons per hour. To be able to do this, the amount of limestone charged at the start of the heat would possibly have to be increased, but the study indicated that this probably would have a negligible effect upon tons per hour.

The indicated potential increase in production rate was of sufficient size to be highly attractive to supervision who were readily induced to attempt the recommended changes in practice which might accomplish this improvement.

It must be noted that much of what we learned from the correlation was not news to anyone familiar with the process. The value of the study lay in the numerical approximations and estimates of the relationships, by which the effect of changes could be predicted quantitatively. These numerical relationships from the multiple correlation were free from the effect of variations of other included variables, to say nothing of personal opinions. In this manner, any increased cost affecting the proposed alterations in the independent variables could be evaluated against the advantages of the increased tons per hour which might be expected to result.

FOLLOW-UP STUDY OF SUGGESTIONS

Since a correlation measures associations without regard to their cause-effect quality, it should be checked in practice. It has little value if its indications can not be so checked to prove the validity of the cause-effect assumptions. The justification of the correlation equation is its successful prediction of the effects of changes in practice.

Such a check was made by comparing the actual change of average tons per hour, over a four-month period following the meetings with open hearth personnel, with the predicted change. The predicted change was calculated by substituting in the estimating equation the actual values obtained during this subsequent period for those independent variables on which data were routinely obtained. Data for 177 heats were collected from routine heat records. Unfortunately, data for some variables used in the equation are not routinely reported, so a complete check of the equation is not available. Grade distribution, average tons tapped, and average non-productive time (delays and bottom time) remained approximately the same during both periods. Other unmeasured variables have to be assumed not to have changed so as to appreciably affect tons per hour.

Certain of the suggestions resulting from the interpretations of the correlation were followed in part, while others were modified due to existing shop conditions. Improvement to the extent of .70 tons per hour resulted. While this falls somewhat short of the predicted increase of 0.99 tons per hour, the difference has been critically examined and is felt to be reasonably explainable. For example, there is the known non-linear nature of the scrap charging time regression line. Moreover, since data for five of the ten independent variables in the equation are not routinely reported, the estimate may be considered to be reasonably close, and the assumptions of cause-effect relationship appear to have been sufficiently accurate for practical use.

Credit for improvement is not claimed by the Quality Control Department. Its function is to interpret the data to the open hearth people in such a way that they may take whatever action they can to benefit themselves. It also had the means of measuring the improvement against a standard - the estimating equation.

APPENDIX

TABLE NO. 1

OPERATING VARIABLES SELECTED FOR INVESTIGATION

Dependent Variables

1. Tons per Hour.
2. Time from Finish Charge Hot Metal to Block

Independent Variables - Group I.

1. Heat Size (Tons).
2. Time from Start Charge to Finish Charge Scrap (Hours).
3. Length of Lime Boil Period (Hours).
4. Time from Tap to Start Charge of Following Heat.
5. Per Cent of Scrap in Initial Charge.
6. Average Weight of Scrap per Buggy.
7. Weight of Limestone in Initial Charge.
8. Boxes of Burned Lime Added During Heat.
9. Per Cent Hot Metal Charged.
10. Per Cent Silicon in Hot Metal.
11. Carbon Drop from End Lime Boil to Block.

Independent Variables - Group II.

1. Fuel Rate from Start Charge to Charge of First Hot Metal.
2. Fuel Rate from Charge First Hot Metal to Start Lime Boil.
3. Fuel Rate During Lime Boil Period.
4. Fuel Rate from End of Lime Boil to Block.
5. Steam Consumption per Gallon of Fuel Oil.
6. Atmospheric Humidity.
7. Carbon Oxidation Rate from First Test to End of Lime Boil.
8. Lime/Silica Ratio of Slag at Block.
9. Per Cent FeO. in Slag at Block.
10. Bath Temperature at Block.

TABLE NO. 2

SUMMARY OF SUMS OF COLUMNS, SUMS OF SQUARES, AND SUMS OF CROSS PRODUCTS - TEN IMPORTANT VARIABLES

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Sum of Col.	6144	4140	5421	6209	4202	1520	523	3803	3712	3213	5404
Sum of Squares	334,582	150,062	334,507	333,301	155,968	34,024	3,609	135,471	113,316	101,797	241,782
Cross Products											
x_0		208,091	259,850	294,718	204,825	60,459	23,445	182,882	182,619	160,581	269,838
x_1			181,342	203,882	139,180	50,373	18,019	129,649	122,770	105,052	180,130
x_2				266,307	186,586	65,360	27,501	150,565	162,781	141,286	237,831
x_3					202,219	80,449	26,019	187,098	186,456	161,505	262,920
x_4						49,230	18,322	131,296	123,805	105,943	184,563
x_5							6,767	45,259	45,807	38,375	63,847
x_6								16,206	15,848	13,546	22,612
x_7									111,741	96,018	165,121
x_8										96,037	160,632
x_9											138,664

TABLE NO. 3

EXAMPLE OF TABULAR FORM FOR COMPUTING THE SET OF SIMPLE CORRELATION COEFFICIENTS (r's)

Col. 1 Variables x_a - x_b	Col. 2 N	Col. 3 $\Sigma x_a \cdot x_b$	Col. 4 Σx_a	Col. 5 Σx_b	Col. 6 $N\Sigma x_a \cdot x_b - \Sigma x_a \cdot \Sigma x_b$	Col. 7 L_a	Col. 8 L_b	Col. 9 $L_a \cdot L_b$	Col. 10 r_{ab}
x_o - x_1	125	208,091	6144	4140	+ 575,215	2018.42	1272.08	2,567,592	+ .2240
x_o - x_2	125	259,850	6144	5421	- 825,374	2018.42	3525.07	7,115,072	- .1160
x_o - x_3	125	294,718	6144	6209	- 1,308,346	2018.42	1763.78	3,560,049	- .3675
x_o - x_4	125	204,825	6144	4202	- 213,963	2018.42	1356.16	2,737,300	- .0782
x_o - x_5	125	60,459	6144	1520	- 1,781,505	2018.42	1393.77	2,813,213	- .6333
x_o - x_6	125	23,445	6144	523	- 282,687	2018.42	421.42	850,603	- .3323
x_o - x_7	125	182,882	6144	3803	- 505,382	2018.42	1571.96	3,172,876	- .1593
x_o - x_8	125	182,619	6144	3712	+ 20,847	2018.42	620.93	1,253,298	+ .0166
x_o - x_9	125	160,581	6144	3213	+ 331,953	2018.42	1549.60	3,127,744	+ .1061
x_o - x_{10}	125	269,838	6144	5404	+ 527,574	2018.42	1009.76	2,038,120	+ .2588
Sum & Check	125	1,847,308	6144	38,147	- 3,461,668	2018.42	14,484.53	29,235,867	
x_1 - x_2	125	181,342	4140	5421	+ 224,810	1272.08	3525.07	4,484,171	+ .0501
x_1 - x_3	125	203,882	4140	6209	- 220,010	1272.08	1763.78	2,243,669	- .0981
etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.

The simple correlation r_{ab} between any two variables, X_a and X_b, is computed from the formula:

$$r_{ab} = \frac{N\Sigma x_a \cdot x_b - \Sigma x_a \cdot \Sigma x_b}{\sqrt{N\Sigma x_a^2 - (\Sigma x_a)^2}\sqrt{N\Sigma x_b^2 - (\Sigma x_b)^2}}$$

The Quality Control Department computes L_a and L_b on a desk calculator.

$$L_a = \sqrt{N\Sigma x_a^2 - (\Sigma x_a)^2}$$

$$L_b = \sqrt{N\Sigma x_b^2 - (\Sigma x_b)^2}$$

TABLE NO. 4

SIMPLE CORRELATION COEFFICIENTS (r's) FOR THE TEN INDEPENDENT VARIABLES SELECTED AS MOST IMPORTANT.

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_o
X_1	+ .0501	- .0981	+ .0007	+ .0022	+ .1626	+ .2309	- .0271	- .0864	+ .1119	+ .2240
X_2		- .0596	+ .1138	- .0142	+ .4055	- .3240	+ .1027	+ .0445	+ .1219	- .1160
X_3			- .3398	+ .2516	+ .0068	- .0814	+ .2367	+ .0873	- .3865	- .3675
X_4				- .1234	+ .1620	+ .2025	- .1451	- .1228	+ .2649	- .0782
X_5					+ .0867	- .0562	+ .0966	- .0402	- .1657	- .6333
X_6						+ .0555	+ .1514	+ .0196	+ .0004	- .3323
X_7							- .1528	- .0890	+ .0559	- .1593
X_8								+ .0810	+ .0309	+ .0166
X_9									- .0192	+ .1061
X_{10}										+ .2588

TABLE NO. 5

SUMMARY OF IMPORTANT STATISTICS DERIVED FROM CORRELATION STUDY OF THE TEN MOST IMPORTANT INDEPENDENT VARIABLES

Variable	Units	Code	$\bar{X}$	σ
X_0 = Tons per Hour	T/Hr.	x=10X-70	11.92	1.615
X_1 = Heat Size	Tons	x=X-150	183	10.18
X_2 = Carbon Drop from End Lime Boil to Block	%	x= 100X	.434	0.282
X_3 = Time from Start Charge to Finish Charge Scrap	Hrs.	x= 10X	4.97	1.411
X_4 = % Scrap in Initial Charge	%	x=X	33.62	10.85
X_5 = Time Between Heats + Delays	Hrs.	x=6X	2.03	1.858
X_6 = Burned Lime Added	Boxes	x=X	4.18	3.371
X_7 =Time of Lime Boil	Hrs.	x=12X	2.54	1.048
X_8 = % Hot Metal	%	x=X-10	39.70	4.967
X_9 = Carbon Oxidation Rate from 1st Test to End Lime Boil	%	x =100X	.257	0.124
X = Oil Rate from Charge 1st Hot Metal to Start Lime Boil	Gal./Hr.	x=$\frac{1}{5}$X-40	416.	40.39

Variable	r	Beta from Computer	Beta from Doolittle Sol.	t-Test	B
X_1	+ .2240	+ .3141	+ .3139	5.98	+ .0499
X_2	- .1160	- .1563	- .1560	2.61	- .8951
X_3	- .3675	- .2676	- .2677	4.58	- .3063
X_4	- .0782	- .1179	- .1179	2.10	- .0175
X_5	- .6333	- .5704	- .5703	11.03	- .4958
X_6	- .3323	- .2557	- .2558	4.45	- .1225
X_7	- .1593	- .2749	- .2747	4.80	- .4236
X_8	+ .0166	+ .1285	+ .1286	2.43	+ .0418
X_9	+ .1061	+ .0981	+ .0980	1.95	+1.2777
X_{10}	+ .2588	+ .0694	+ .0893	1.61	+.00357

(Table Continued on Next Page)

TABLE NO. 5 (Cont'd)

Significance Levels of "t"

R^2 = .7217
R = .8495

t.01 = 2.62
t.05 = 1.98
t.10 = 1.66

S_{x_o} = 0.852 (Standard Error of Estimate of X_o)

Estimating Equation:

$$X_o' = + .0499\ X_1 - .8951\ X_2 - .3063\ X_3 - .0175\ X_4 - .4958\ X_5 - .1225\ X_6 - .4236\ X_7 + .0418\ X_8 + 1.2777\ X_9 + .00357\ X_{10} + 4.41$$

NOTE: $B_a = \frac{\sigma_{x_o}}{\sigma_{x_a}} \beta_a$

BIBLIOGRAPHY

(1) Journal of Applied Physics, Vol. 17, No. 4, pp. 262-272, April 1946. "A Computer for Solving Linear Simultaneous Equations."

Electrical Computer Operation Manual, Consolidated Engineering Corp.

Annals of Mathematical Statistics, Vol. XVI, No. 4, Dec. 1945. "A Criterion of Convergence for the Classical Iterative Method of Solving Linear Simultaneous Equations," by Clifford E. Berry.

(2) Statistical Methods for Research Workers, R. A. Fisher, Hafner, 1950, Page 165.

ATTAINING UNIFORMITY IN PROCESSING RUBBER BATCHES

P. E. Gnaedinger
Dominion Rubber Company Ltd.

I - INTRODUCTION

The attainment of uniformity in batch processing can be assisted by the use of operator control charts. In fact it is doubtful if best results can be attained without them. The influence that they may have is largely dependent upon the philosophy that guides their administration. There are four aspects that will be considered under this head. First is the concept that a so-called batch process should be treated as a continuous process. Second is the predominant importance of the human element in process control. Third is the conviction that the majority of workers are inherently interested in doing a good job. Fourth is that it pays to expend effort to make process control data easy to interpret. These four points will be discussed in this paper against a background of rubber batch processing through the stages outlined below.

The processing begins with masticating the crude rubbers to bring them to a consistency suitable for the subsequent stages of processing. The next step is weighing out the various chemicals and rubbers into batch lots prescribed by the various formulae. In the next stage the materials are combined together in rubber mixing mills to form a homogeneous mass that is slabbed out and cooled ready for calendering, extruding or other conversion processes. The compound weighing and mixing operations comprise the batch processing part of the rubber manufacturing process. They are the heart of the process. How well they are done determines in large measure the quality of the finished product that issues from the other end of the plant.

The important variables are weight, time and temperature. Sometimes some of the weighing is mechanically controlled, and sometimes the mixing time cycles, but generally the important factors are under the direct control of the workers.

Quality is gauged by taking a sample of each completed batch and subjecting it to various qualification tests before it is released for further processing. Some of the tests are indirect in that they do not evaluate the characteristics that are of concern. This is an inherent difficulty in control and exists because of the very nature of rubber.

Thus, there are two good reasons for applying control chart techniques to rubber mixing, first the reliance that must be placed on the human factor in the process and second, the uncertainty that surrounds the test data. There is also a third reason that bears considerable weight - it is better to know that processing is being done right while it is being done, rather than wait until the work has been completed before trying to find out if it has been done correctly. The third reason grows in stature in the light of the second, for the means of detecting potential trouble in subsequent processing is certainly not too good with compounded rubber stocks. The penalties from lack of control can be severe.

II THE BASIS FOR QUALITY CONTROL

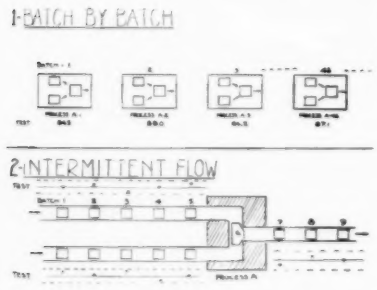
II - 1 Batch or Continuous Process

The mixing of rubber stocks is called a batch process because the raw rubbers and chemicals are weighed out in batches and combined together in batch lots. This processing method usually leads to thinking of the process in a batch-by-batch sense. That is, each batch is treated as a separate lot without any relation to those that have gone before, nor of those yet to come. Each batch that is tested is judged upon its own merits. It stands or falls as an individual.

Actually the chief reason why batch production prevails is because of physical limitations of the equipment. If the stock could be mixed continuously, it would be.

The fact that the material goes through the process in batches should not prevent us from thinking of the process as being continuous. Each series of similar batches is produced, or is expected to be produced under the same system of conditions. The equipment should operate consistently, the men should follow standard procedures, and the material should come from a homogeneous source. Under these conditions, the output should be homogeneous although produced intermittently.

Under this concept each batch becomes a sample from a continuous run. Results from each test sample should be viewed, not as isolated tests, but more in their relation to preceding tests. The information conveyed by each test is supported by the accumulated knowledge of the process as built up by experience.



(Fig. 1)

Fig. 1 illustrates the difference between the two concepts. This is a simplified case where two materials are being combined in a process to form a third material.

In Fig. 1-1 each batch is judged on the basis of its individual test with respect to the specified quality. Each batch stands upon its own merits.

In Fig. 1-2, the successive batches are pictured as being fed in an intermittent stream to the process, and leaving it in the same way. Control charts present the test data. By this means the chart establishes a pattern of the variation to be expected from batch to batch. Often the variation can be attributed largely to the error of sampling and test procedure.

With this technique each test becomes an indicator of whether the material is continuing to be about the same as in the past or whether something has happened that makes it significantly different from other batches of the same type.

The implications behind this manner of judging quality are quite different to the ones that arise when a test is used more as an attempt to pin an exact measure of quality on a specific lot of material. If the batches are treated as individuals the attention is on each batch, the prime consideration being what to do with it if it is off-standard. But when production is treated as a continuous stream, attention is focused on the process and its variables. Emphasis is shifted. The aim becomes to prevent the variables from getting out of control.

The distinction between the two concepts is perhaps subtle but it makes a great difference in the long range results obtained. When attention is directed to the variables in the process, it will not be long before steps are taken to keep the variables under control. That is where true quality control begins to develop. For 'control' implies the prevention of trouble, or detecting it in the early stages before much damage can be done.

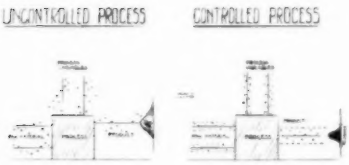


Figure 2 developed by A.E.R. Westman illustrates this principle in an interesting way. In the left hand figure the variables are left on their own. The finished product quality is erratic. In the right hand figure the variables have been brought under control as evidenced by the control charts. Finished product quality is more consistent and is within the tolerance limits.

(Fig. 2)

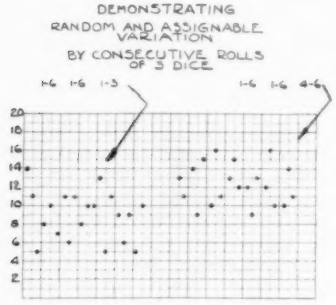
There is a by-product of such quality assurance. When it is known that process variables are under close control, it is safe to predict with confidence the quality of production yet unmade. Under these conditions the number of samples taken and the amount of testing done can sometimes be reduced considerably.

II - 2 The Human Element.

At this point it seems appropriate to make a reference to the hackneyed elements of variability, men, machines and materials. The reason is to demonstrate just how these elements do affect variability. For this purpose three dice can be used, two numbered in the conventional way 1-6 and the third 1-3. Each die represents one of the elements, men, material or equipment. Now if these dice are rolled a number of times and the total points plotted in sequence on a chart to a suitable scale, the resulting pattern will be typical of product quality variations. These arise from the variations caused by interaction of the three elements of variation, men, machines and materials.

Next, change the die numbered 1-3 to one numbered 4-6 and assume this represents either a change of operator, or a machine adjustment, or a new lot of material. In other words, a change has been introduced to one of the variables of the process, say for example, a

different operator is now working on the job. Next make a new series of rolls with these dice and plot on the chart as a continuation of the previous series.

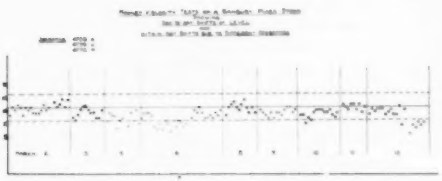


(Fig. 3)

Figure 3 is a sample of how the two series would look. Both series exhibit similar variability, but there has been a shift of the average quality level.

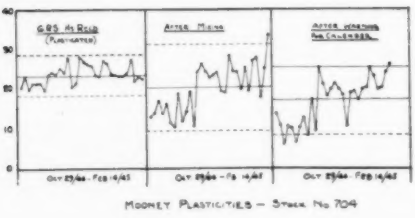
Such shifts take place in actual practice as demonstrated in Fig. 4. In this case changes in quality level from day to day were detected by the control chart and traced to changes of operator. The way to improve uniformity, quickly became obvious; it involved training each of the operators to work in a standardized way and then establishing controls to see that they did.

Before studies of this kind were started, it had been customary to accept such variability as being inevitable just because rubber is highly variable as received in its raw state.



(Fig. 4)

While it is true that men, machines and material are the broad bases for classification of process variables, we must be careful to blame the right base when trouble arises. For example, in the case of Fig. 4 the man who uses the stock might attribute his troubles to the variability of the material he is using. In a sense he would be right, but basically his material would be more uniform if the men who mix it in the first place would all follow the same techniques.



(Fig. 5)

Fig. 5 illustrates the same point. In this case the control chart for the raw material as received showed normal variation in quality. However after it had passed through the mixing operation, a shift of level appeared at the thirteenth test and the same shift persisted through the next stage of the process. An analysis of this kind emphasizes the fact that often our processing troubles stem from the things we do or don't do to the materials after they have started through process, rather than from the materials themselves.

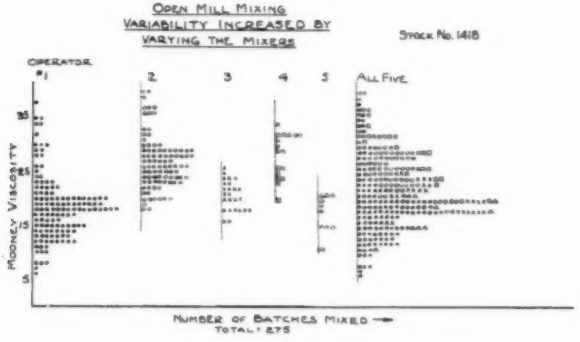
Most of these errors of commission and omission can be traced to the unpredictable human behaviour. For example if some unsuitable material gets into process, should we blame the material or rather the man who accepted or released it? If a machine has produced off-standard material, should we blame the machine or the man who operated it or perhaps authorized its use?

Candid answers to questions of this kind will probably lead to the conclusion that the predominant cause of variability is really the men in the plant. Generally speaking, the quality variations introduced by the operators consciously or otherwise are much more difficult to foretell than those arising from the materials or equipment. From this, it is apparent that the control of workmanship is usually the key problem in quality control. These observations are not made as a criticism of the man as an operator, but only to localize the problem.

IJ - 3 Pride of Workmanship

The way to get consistent good workmanship from a group of people starts with a genuine conviction that most workers have a natural desire to do things well. With this as a foundation the next thing is to find effective ways to encourage each one to do his best. The first essential is to systemmatically evaluate the quality of each man's work, and second to make his rating a personal matter with him.

Unfortunately our industrial system does not work that way. One of its handicaps lies in the fact that relatively few of the individuals engaged in making a product ever see it when it is finished unless they happen to buy it for their own use. Certainly there are a great many of them who never see the direct result on the finished product if they deviate from specified instructions. Usually it is not practical to keep the operators informed of these things. Thus the lack of quality consciousness is quite widespread.

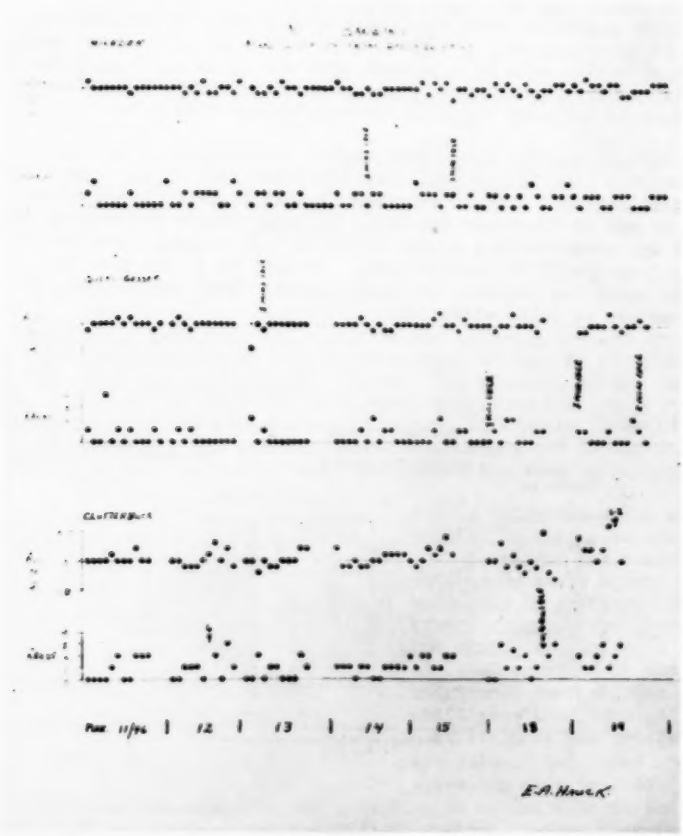


(Fig. 6)

Management is largely to blame for the situation. That means everyone from the supervisor up. Often an operator does not know exactly what is required, or he may not have the skill. Frequently no person checks him when he is off the beam and more commonly still, no one praises him when he does a good job consistently.

Process control records contain plenty of data that can be turned to good account if it is personalized. Fig. 4 is taken from a typical control chart of routine viscosity tests that are made on every batch mixed. The charts are kept in the quality control department. If the variability of a stock seems to be excessive, further analysis is made to try to find the cause. Such an analysis shown in Fig. 6 should be discussed with the operators to emphasize the necessity for following instructions closely and consistently.

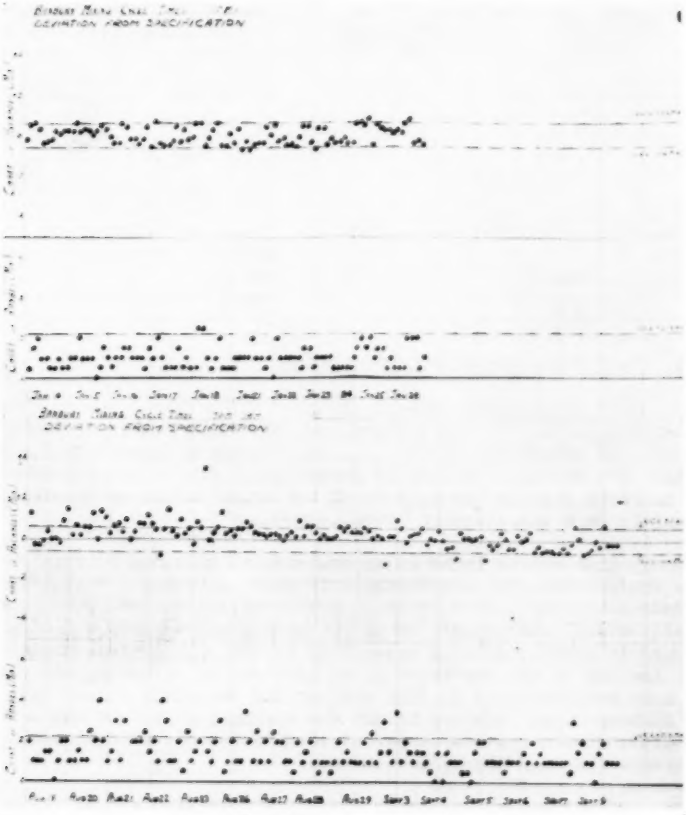
Fig. 7 is an operator control chart on batch mixing cycle times. It shows comparative performance of three shift operators on the same machine. The record is an average and range chart for each five con-



(Fig. 7)

secutive batches, the average being the average deviation from the specified mixing time, and the range the difference between the longest and the shortest.

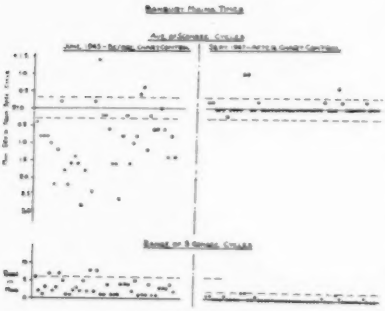
These charts are posted beside the work station. They bring home to the operator the fact that his work is being watched closely. The effect is similar to a ball player's published averages. Pride of workmanship is stimulated because the rating is based upon factors that are readily understood, and they are directly within the man's own sphere of influence.



(Fig. 8)

Best results come from trying to encourage the good worker rather than scold the poor one although sometimes the latter becomes necessary. Very often poor work is neither willful nor due to carelessness. It may be beyond the operator's control. Fig. 8 is an example. The average and range of cycle times in the lower section of the chart are not as good as those in the upper section. The former are for the month of August when the weather was hot. The upper chart is for January

when the weather was cold. The more erratic performance during August was due to difficulties encountered by the men, because the temperature of the cooling water was too high for satisfactory processing. When facts like this become known to Management, it enables them to take proper corrective action, and sometimes forestalls laying the blame in quarters where it is not deserved. Such misguided actions discourage interest.



(Fig. 9)

The value of control charts lies in their sensitivity to trouble. They are of little use, however, unless Management acts promptly on the things that are bound to come to light. Intelligent interest on the part of Management is a powerful influence towards stimulating an operator's pride in his work. On the other hand, the improvement in uniformity of mixing cycles shown in Fig. 9 could not possibly be attained without the sincere interest of the man on the job. These two factors must be coupled together to assure success.

II - 4 The Effectiveness of Control Work.

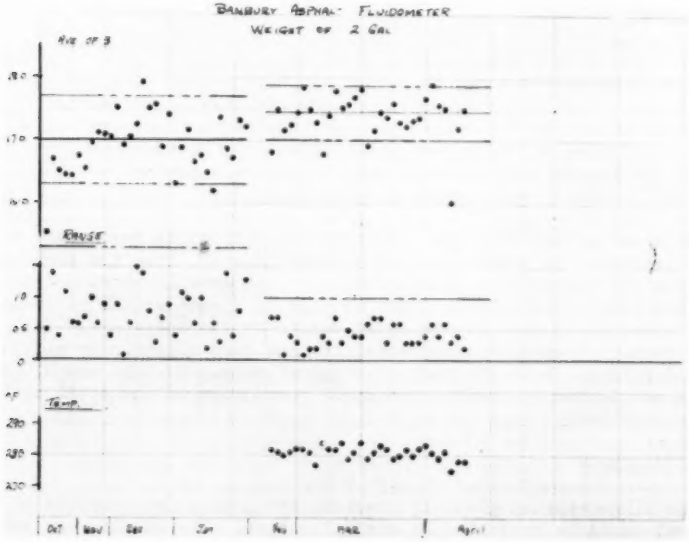
In every organization engaged in production there is a vast number of examinations made on the process, and materials being processed in the course of a day, every day. All this checking taken together forms the structure by which quality is controlled. How it is used determines in large measure, the success of the organization in maintaining its quality on a satisfactory level.

Every time anyone looks at an instrument, examines material, or checks the process, you are buying information. Much of this investment does not yield a fair return. A minimum is derived from the multitude of casual checks made by people throughout the day and of which no record is taken. In many cases this is the only control exercised, yet at the end of the day there is no information regarding what deviations have occurred, nor in fact whether the necessary checks have actually been made. Whether or not the appropriate action has been taken often does not become apparent until after the product has been completed - but then it may be too late.

The effectiveness of general quality supervision can be improved by putting it on a systemmatic basis. This is accomplished by organizing a routine of spot inspections to be made at regular intervals. The frequency may be tied to production or it may be on a time basis, running all the way from once per half hour to once per day or longer. A record should be kept of the observations. Usually a control chart will be of value to provide continuity to the history and to emphasize unsatisfactory conditions.

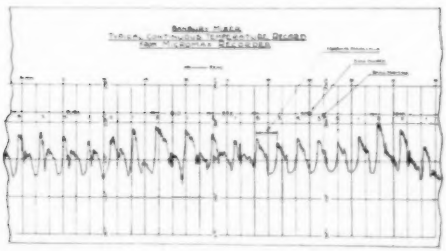
This idea has been used successfully to improve the accuracy of compound weighing. It is now standard procedure to spot check the

batch weights of chemicals and rubbers prior to mixing. With certain critical stocks every batch is checked. With others, the frequency is reduced to several checks per shift. The time at which they are made is random. Here again the operator's interest has been aroused by means of control charts thay show deviations from standard weights. This procedure takes but a few hours each day, and the cost represents a small fraction of the trouble it has eliminated.



(Fig. 10)

Fig. 10 shows a series of spot checks taken on an automatic metering device. The successive points show variations in the weight of 2 gallons of liquid asphalt. The second half of the chart shows improvement gained by better temperature control. Spot checks like this made daily and plotted on a control chart are a great help in giving quality assurance.



(Fig. 11)

Fig. 11 is a continuous temperature record from the mixing operation. Charts of this type are usually scrutinized periodically and then filed in a drawer, provided maximum specified temperatures have not been exceeded. That is typical batch psychology, but continuity and uniformity of processing are important factors also, and these charts are very difficult to analyse from this point of view. That is particularly true

with large variety and short runs, when frequently some stocks do not repeat before several days or maybe weeks have passed. Consequently, temperatures are read off the original chart at three important stages of each cycle, and replotted on control charts that are maintained for each of the critical stocks. By plotting averages and range of five consecutive batches, it is possible to condense the data to accomodate a year's production on a single chart. It is a simple matter to maintain an individual chart for each stock, and by this means, quality is interpreted as a running record rather than as a batch by batch acceptance.

The temperature charts serve another purpose. It will be noted that each hump on the chart is a complete batch cycle time and provides a convenient means for checking the operator's conformance to the specified cycle time. For this purpose, the batch times are scaled off in groups of five and the average deviation from standard and the range between the longest and shortest are plotted on a control chart. Figs. 5, 7, 8 and 9 were compiled in this way.

It might appear that the time spent in replotting instrument charts is unwarranted. It should be kept in mind however, that the work can be done by a clerk, and once it has been done, it greatly shortens time and clarifies interpretations that may be made by higher priced personnel. The information is also comprehended better by the operators. The work can be kept at a minimum after stability has been attained by applying sampling principles to the mass of original data. By this means, control is maintained by analysing a small percentage of the total available data on hand.

III CONCLUSION

The illustrations given in the text are just a few examples of many that could be presented of a similar type. The charts were selected from earlier periods, because that was the most interesting stage when progress was sometimes spectacular.

The plant they were taken from uses 35 different kinds of rubber, and about 300 different chemicals. These are combined to form over 1000 different compounds of varying physical characteristics. Currently there are over 3000 control charts in active use in the compound weighing and mixing departments alone. These cover the routine and special testing of colour, specific gravity, hardness, viscosity and the physical characteristics. In addition to this, there are operator control charts posted beside every operation to indicate the individual quality performance of each man in the department in respect to some important phase of his work.

The charts have become an indispensable tool to the operating and technical staff. Of greater importance is the service they are performing in creating operator interest in the quality aspects of production. These factors together with the underlying philosophy as discussed in this paper, have combined to accomplish a substantial decrease in off-standard batches during the past six years. During that time, the percentage of suspected batches held because of failure to meet qualifying tests has decreased by over seventy-five percent to reach the best level ever experienced by the plant.

CONTROL CHARTS AND QUALITY LEVEL APPLICATIONS IN TEXTILES *

Edward S. Rudnick
Wamsutta Mills

As in most cotton mills, quality control at Wamsutta breaks itself down into four classes:

a. Control of the raw cotton by the Cotton Department
b. In-process controls by roving departmental quality inspectors
c. Cloth Room and finished good 100% inspection
d. Laboratory controls (embracing all of these)

Since only the last of these classes deals with statistical quality control techniques as the Q.C. Engineer knows them, this paper will outline what Wamsutta's Quality Control Department has done first, with control charts and second, with quality level determinations.

As most of us know, the main function of a control chart is to separate chance variation from variation due to a significant change in the process being investigated. When averages of two or more test results are plotted on a time basis, the charts become a much more sensitive measure of such significant process changes -- the lines which separate the chance variability from the "out of control" situation are much closer together than they would be for individual points. We also realize that by plotting such points in a time sequence on a so-called $\bar{X}$-Chart, significant process trends can be noted even though all might be within limits.

Similarly, by plotting the spread or range represented by the individual data making up each average, we have a picture of the variability of the process and can note, by the range control limits, when this exceeds the normally expected variability. In each case, the control limits are based on the past performance of the particular process.

Before diving head first into the use of control charts on the various textile processes, it's a good idea to make some preliminary studies to find where and how they can most effectively be employed. The "where" question usually resolves itself to the key processes of the particular mill. However the statistical tool of analysis of variance is quite helpful in determining which factors must significantly influence the variability of each process, so that these can be taken into account in setting up a control program for the various operations. This will not be covered here as it has been more fully discussed in previous papers by the speaker. (1)

---

* Some portions of this paper are from a similar paper delivered by the speaker at the Textile Division, A.S.Q.C. Meeting, Sept. 1952, and at the Sixth New England Q.C. Conference.

The "how" question on the use of textile process control charts can be partially answered by following the steps of Wamsutta's program.

After it had been decided, with the help of the analyses of variance, which processes were to be controlled, it was necessary to determine what sample size to use for each process. This was done by computing the average coefficients of variation both from past records and from the data of the general analyses. Tables such as those recommended by Committee D-13 of the A.S.T.M.[2] were used. To determine which "allowable random sampling error" to use, the overseers concerned were asked whether they would accept any conclusions drawn as being correct 19 out of 20 times (or whatever other probability they desired).

Forms were then made up for the collection of data to comprise the populations on which the quality was to be based. In some cases past data were used for this purpose.

In the case of the yarn, the items controlled were the mean size, the range of the size, the mean strength, and the range of the strength for each count each day. For simplicity of calculations, ranges were used instead of standard deviations.

These yarn charts are based on the English System of control limits rather than the American "three sigma" system. Inner control limits are used at the 2.5% level and outer control limits at the 0.1% level.[3] The area between the two sets of limits is called the "Warning Region". Thus although five times in each 200 a point will fall in the "Warning Region" due to chance alone, the probability of two successive such points being due to chance is slightly greater than 6 times in 10,000. Using the straight three sigma limits, such points would probably be considered "in control".

For the roving process, it was decided to use four bobbins from each frame each day, while using one average and range chart for each type of roving. Thus if on one day there are three frames on a single type roving, then the sample size is twelve. Accordingly there are three sets of control limits on each of these average and range charts (in different colored pencil) to indicate the limits for sample sizes of four, eight, and twelve. In the case of the $\bar{X}$-Charts, the narrowest limits are for the largest sample size, while the reverse is true on the range charts.

Almost from the beginning the control charts were warmly received. The superintendents and overseers observe the trends periodically. But the controls do not stop here. Daily reports interpreting the charts are distributed to the general manager, the superintendents, and the overseers. The factors being controlled for each count of yarn are reported in columnar form as, "Weak, Out of Control", "Warning Region Light", etc. When five points in succession fall on the same side of the mean, where the probability is 1/32 that this is

due to chance alone, this is reported as "Continues Heavy, Within Limits"; "Continues High-Range, Within Limits"; etc. Also, weekly reports are made for each item, giving the percent of "Good", "Warning Region", and "Out of Control" points so that the overall quality records of a department can be compared from week to week.

The material value of such reports is obvious, particularly if all "Out of Control" places are circled in red and if the reports are distributed the same day that the samples are taken. However, the psychological value is also great, since the overseers know that the men above them are reading the same report. The results manifested themselves in the following ways: almost all the ranges showed consistent downward trends; the weekly reports showed definite increases in the percent "Good" and decreases in "Out of Control"; ends down in spinning and weaving decreased; experienced mill men found themselves speaking in terms of ranges; many investigations were made to determine why certain factors had gone out of control; and everyone became more quality-conscious.

This method of controlling the quality has been made flexible for easy revision. A running record of the daily means and ranges is kept so that every few months the charts can be revised and the limits narrowed to fit the new quality standards.

Besides the process control charts on Yarn and Roving, Wamsutta also uses "$\bar{X}$" and "R" Charts on the Finisher Drawing and Ribbon Lap processes -- Wamsutta being almost a 100% combed cotton mill. For the first year or two of this program, all the testing was done in the Laboratory and all the charts maintained in loose leaf fashion on a large slanted table top. A system of cross reports was installed whereby each production gear change was reported to the Lab. and these noted on their respective charts in the form of arrows up or down to denote "assignable causes of variation".

Then certain modifications were made. After a couple of revisions of the roving it was noted that this process ran almost continually in control with very rarely any need for gear changing. Instead of further narrowing of the control limits, it was decided to reduce the frequency of testing from every day to twice a week.

To get more rapid and closer control of the two key processes -- ribbon lap and drawing -- the testing was taken out of the Laboratory and back into the mill. Now these machines are sized twice each shift and the averages and ranges immediately plotted on a separate control chart for each machine. Only when the point falls beyond the control limit (as verified by an immediate check size) is the gear on that machine changed. The Quality Control Department supplies the charts with the limits already on them. Since the time this system was installed, the number of gear changes on these processes has actually been reduced and yet the product on the subsequent processes is more even.

At present, besides the Spinning Mill processes, $\bar{X}$ and R Charts are maintained for different styles of greige and finished cloth strengths, for departmental waste records, and even on the accident frequency charts of Wamsutta's Safety Department. "P" Charts are maintained on the Percent Seconds found by each group of inspectors in the Sheet Factory.

The second part of this paper deals with Wamsutta's use of quality level determinations by our Quality Control Laboratory. Since we have not as yet found a way of eliminating 100% inspection of our greige and finished fabrics, and completed sheets, towels, and other unit items, we might loosely define these quality level determinations as a reinspection on a sampling basis.

Actually, it consists of a very critical sample inspection for all (even the most minor) of visual defects. It is a system first introduced by Mr. A. G. Klock, of Bigelow-Sanford Carpet Co., except that at Wamsutta we have refined the system by weighting the defects according to their seriousness. The main purpose of any such system is to express quality numerically (in this case in terms of visual defects). Once quality is expressed in numbers (by going from an "attribute" system to a "variable" system), we can treat it in different ways, ie., take averages, set up control charts with statistical limits, observe trends, subdivide, etc.

At the end of this paper is a copy of the "score card" we use for the quality level determinations of our finished sheets after they have been subjected to 100% inspection. One of these score cards is used for each item being inspected -- in this case sheets ready for packaging. Every possible visual defect, both major and minor, has been listed on the left and broken down, as well as possible, for departmental source.

There are four possible degrees of seriousness for each defect which have been pre-determined for the respective faults. For instance, one fourth degree slub would be the equivalent of four first degree minor slubs. In the next to the last column, a weight is given to each type of defect since a first degree Alabama measle (tar speck) is not as serious as a first degree drag-in, which in turn is not as serious as a first degree broken end. Similarly the rest of the weights have been more or less arbitrarily set up as they effect the overall sheet quality.

In practice the quality level inspection works like this in our Sheet Factory. One of the Q.C. Laboratory girls selects random sheets from the line, just after pressing prior to packaging. These are all supposed to be first quality sheets which the 100% inspectors have passed. On a special large table she spreads the sheet out and then very carefully goes over it for even the most minute of flaws, tallying the various defects on her "score card" for that

sheet.

She then totals her tally marks, counting second degree defects as two, third degree as three, etc. Next, these totals are multiplied by the respective weights to get the "score" in the last column for each defect. Finally, these scores are totalled vertically to get the total score for each departmental source and the grand total for the sheet being inspected.

Since it is possible to have a Second or Third Quality sheet without necessarily a high total score, she notes under "comments" if any of the defects were serious enough to have kept the sheet from being a First and also if the dimensions vary beyond the tolerances. We all realize that 100% inspection can never be 100% efficient. Thus, an important by-product of these quality level determinations is the measurement of how perfect our 100% inspection is. Since the inspector's number appears on each sheet, it is a simple matter to go back to the inspector and show her the type of Second Quality material she has passed. As in many other Q.C. applications, this is a very important psychological tool, especially where records are kept of which inspectors passed "bad" material and warnings are issued for repeated violations.

The quality level inspection of each sheet is so critical that it takes several minutes to inspect one sheet. Even though we make what is considered to be the finest quality sheet, we have yet to fine one scoring zero, after about 10,000 such sheets have been inspected. Without such a critical inspection as this, much of the effectiveness of such quality level determinations would be lost.

At the end of each week, after about 100 sheets have been sampled and inspected, the "score cards" are summarized to get the average score for each department (ie., the cotton defects score per sheet, spinning defects score per sheet, etc.) and the overall average score per sheet. Each of these figures is reported to those supervisors concerned in the form of a weekly report comparing them with the average overall departmental and total scores for all previous weeks. They are also plotted to observe trends in each department's defects contributing to the overall sheet defects. When the defects from a certain source show a rising trend the individual score cards can be studied to see which flaws are most prevalent and these can be stressed for improvement.

The total scores of the hundred or so sheets inspected for the week form a "C" distribution of "defects per unit". The standard deviation of such a "C" distribution is merely the square root of $\bar{C}$, the overall average score per sheet. Thus, hypothetical control limits are figured each week at $\bar{C} \pm 3\sqrt{\bar{C}}$. All score cards falling above or below this amount are picked out and observed, since these are the exceptions to the chance variation rule -- particularly when they exceed 1% of the total. In this same weekly report, these "out of

control" sheets are listed with the reasons for extra high (or low) scores noted.

Finally, on this weekly Sheet Quality Level Report, any Seconds or Thirds are listed with the inspector's numbers who let them get by; they are expressed as a percentage of the total inspected and compared with the respective average percentages for all previous weeks. Thus it is a simple matter to compute the efficiency of the 100% inspection at any time. For instance, if the inspectors have been rejecting 8% of the sheets they inspect as Seconds and if the sampling inspection reveals 2% Seconds getting by with Firsts, then the efficiency of the original inspection is only about 80%. Fortunately, our definitions of First Quality are so high that even if 2% were to get by, we would still be in no real trouble.

Occasionally, quality level determinations are made of Second Quality sheets to determine which defects are most prevalent. Records are kept of the number of Firsts being rejected as Seconds and the number of Thirds getting by with the Seconds.

Without going into the details here, I can say that in similar manners, quality level determinations in the form of sampling inspection, are made on our pillowcases, towels, and piece goods. The main problem in the latter case is breaking down yardages to standard units. We solve it by selecting random rolls and observing the first ten yards of each roll -- ten yards being the unit of measure. Thus, the quality level report for piece goods is in the form of a two-dimensional table showing both the defect breakdowns for each style and for each source of defects.

In somewhat of a summary fashion, I have tried to outline here two aspects of Wamsutta's formal Q. C. program -- first, the use of control charts and second, the use of quality level determinations.

---

(1) "Statistical Quality Control at Work in a Cotton Mill". Textile Research Journal, Vol. XX, No. 10 (Oct. 1950)

(2) Committee D-13, A.S.T.M. Standards on Textile Materials -- Appendix V, Proposed Recommended Practice for Calculating Number of Tests, Philadelphia, 1949.

(3) Brownlee, "Industrial Experimentation", Brooklyn, Chemical Publishing Company. 1948.

TESTING PRODUCTION DATA FOR NORMALITY

Frank G. Norris
Wheeling Steel Corporation

This discussion considers the following points related to testing production data for normality:

1. Characteristics of Production Data.
2. Possible sources of departures from normality.
3. Reasons for testing.
4. Methods or statistical procedures of testing.
5. Interpretation of results.

1. PRODUCTION DATA

Much of this discussion will possibly be found pertinent to quantitative data from any source regardless of its relation to an industrial production process, and to this extent the term "production" may well have been omitted, making the title to read simply, "Testing Data for Normality". The full title is preferred in order to focus attention upon the broad aspects of production as distinguished from a narrow interpretation of the meaning of Quality or Quality Control that might otherwise be inferred. In general, a Quality Characteristic is regarded as: any feature of a product or process that can be expressed numerically by either count, measurement, or computation. For example, the amount of production per unit time is a quality characteristic of a process and is in the area of the present paper. The subjective interpretation of quality in terms of conformance to a specification, (i.e., someone's needs, either actual or fancied) is only a part of and not the whole meaning of quality. Therefore, the term production data includes industrial data from many sources, but does not necessarily exclude data pertaining to natural products or processes.

2. NORMALITY

The term anormal will be used rather than abnormal to describe significant departures from normality. By significant is meant greater than the differences expected as the result of random sampling. A normal curve is of the type $y = e^{-x^2}$. For the large group of people (including the author) who find it more convenient to use a table rather than to evaluate a definite integral, normality may be defined as conformance to the curve described by the tabled values of either height (ordinate or y distance) or area for various values of abscissa or x distance.

One school of thought infers that a normal pattern is the natural consequence of a large number of observations. This thinking implies that failure to develop a normal pattern is due, at least in part, to insufficient data and can be corrected by accumulating more data.(1) Instances will be mentioned later of the skewed pattern disclosed by a group based on over 100,000 observations. It would seem that gathering additional data is not the next step in a study at this stage.

Many types of data tend to approach a normal pattern. From one point of view it may be said that the normal curve is approached, but possibly never completely attained in a production process or product. The presence of a symetrical unimodal pattern in data from a wide variety of sources has lead to some thinking that the normal pattern is the expected pattern for all normal or usual situations; and that an anormal pattern is abnormal in the sense that it indicates that something is wrong or at least, unusual. In general, a skewed pattern can be expected if the values are subject to some natural limitation. A few examples may clarify this point. The distribution of the sulfur content of basic open hearth steel normally (or so frequently that it is expected) exhibits a skewed or anormal pattern. Moderately high values (above the median) are relatively frequent and extremely high values are not impossible. Low values are limited by the chemical nature of the process and negative values are mathematically impossible and without meaning. A high value of more than twice the mean may occur. A symetrically similar low value would be less than zero and has no physical significance. The same reasoning applies to the FeO content of basic open hearth slag. A floor is imposed by chemical considerations, but there is no corresponding ceiling to the values. The distribution of values for a characteristic such as yield is skewed in the opposite direction. The yield cannot exceed 100%, but it can (and sometimes does) reach quite low values.

Other limitations that lead to an anormal distribution may be related to a physical or functional relation among the characteristics being considered. If a linear dimension is normally distributed, area, or volume cannot be normally distributed. The variation in dimensions of sheets with respect to gage, width and length are close to the normal pattern. Variations in the weight of single sheets follow an anormal pattern. The classic example of this situation is the distribution of weights in a group of ball bearings, having normally distributed diameters.(2)(5)

Another source of departure from the normal pattern is a shift in the process (or product.) This condition is equivalent to the combination in one distribution of information coming from more than one source. If the difference between the sources is small, the combined distribution may depart from normal by having a flat top. As the difference increases, the combined pattern becomes bimodal (having two peaks).

3. REASONS FOR TESTING

This discussion is intended to be suggestive rather than exhaustive. Each investigator will be able to think of other reasons pertinent to his own situation:

3.1 As a preliminary step to further study of the data.

3.11 Test to see if need for more data is indicated.

In certain techniques the instructions are to accumulate data for 50 (or some other definite number) of samples or until a normal pattern develops.

Such a requirement is difficult, if not impossible to interpret or comply with unless supplemented by a test for normality.

3.12 Test for general conformance. Some formulas and procedures are based on the assumption of normality. (6)(7) This assumption is stated in the derivation, but the appropriate test is not always mentioned. Sometimes the limitation is either overlooked or misunderstood by the investigator. In fact, the requirement for conformance to normality, "is a custom more honored in the breach than the observance". It still remains, however, as a definite reason for testing normality.

3.13 Test for conformance of a transformed variable. If the original data are badly and systematically anormal, sometimes a transformation of variable is used to permit study of a normal distribution. A test of the transformed variable for normality is a logical means to determine if the transformation has accomplished the purpose of normalizing the distribution.

3.2 For description of a sample.

If a distribution is normal, it is completely described by the average and standard deviation. A test for normality is helpful in deciding what statistics contain the essential information. (3) An estimate of how well a smoothed normal curve fits the data is considered a part of a description of a sample.

3.3 For prediction.

Closely related to the problems of description is that of prediction. How many values were above (or below) a given value? How many such values may be reasonably expected in future samples? A test for normality forms the basis for some of the decisions in answering the above questions. It cannot be emphasized too strongly that if assignable causes are affecting the data, reliable prediction of future values is impossible. Therefore, in this situation tests for normality must be supplemented by control charts, analysis of runs, or some other method of detecting assignable causes.

4. METHODS OF TESTING

Appropriate methods for the various needs outlined above are well known and described in the literature. No new methods are proposed in this discussion.

The methods that will be illustrated are:

1. Determination of k as an indication of skewness.
1a. A "t" test for the significance of k.
2. Determination of Bowleys measure of skewness.
3. Plot of the cumulative frequency distribution on arithmetic probability paper.
4. Selection of rational sub-groups and a control chart of averages to detect sources of anormality.

Methods are best illustrated by means of data. The source of the data used is such that release is beyond presently available authorization. It can be said that they are intimately related to a quality characteristic of a steel mill production process. Furthermore, they have been found to be typical of the type of data accumulated from many unrelated processes and products. For this reason it is believed that identification of the variate is not essential to the illustration of the method for testing for normality.

A well known estimate of skewness is based upon the cubes of the observations and is called k in the ASTM Manual which designation we shall follow. It is called g by Snedecor and in other texts is variously referred to as α_3 or $\sqrt{\beta_1}$. This measurement, however designated, is the third moment about the average divided by the cube of the standard deviation. k is a ratio without units. The standard error of k depends only upon the number of observations in the distribution being described and is:

$$\sigma_k = 6N\,(N-1)/(N-2)(N+1)(N+2)$$

Applying the t test with infinite degrees of freedom indicates that any value of k greater than 2.58 σ_k discloses (at the .01 level of significance) skewness greater than expected to arise from sampling a normal population. These limiting values of k and those at the .05 level are shown in Table 1.

A measure of skewness that is simple to compute and easy to understand is based upon the relations among the three quartiles:

$$sk = Q_3 - 2Q_2 + Q_1/Q_3 - Q_1$$

This relation is called Bowley's measure of skewness and lies between -1 and +1.(4) It is a linear estimate in the sense that no higher powers such as squares or cubes of the observation are used.

In order to become familiar with the application of these tests and to visualize their interpretation two distributions have been selected for illustrations.

Figure 1 shows two selected distribution of 100 items with values of both k and sk. For Figure 1 the value of k is slightly above the 01% point (.624) which indicates greater skewness than can be attributed to sampling errors from a symetrical population. For Figure 1B the value of k is between the .01 and the .05 point.

Figure 2 shows a distribution of 130,000 observations. Casual inspection of Figure 2 indicates skewness. More details of the departure from normality are disclosed by a plot of the cumulative curve on a distorted scale. The cumulative distribution is computed from the observed distribution. Figure 3 shows this curve plotted on the usual scale. Figure 4 is the same curve plotted with the vertical scale distorted in such a way that cumulative data from a normal distribution fall on a straight line. The slope of the line is an indication of the standard deviation of the group.

The two straight lines on Figure 4 fit the points fairly well, which suggests that the distribution can be represented by portions of two normal curves:

Interval	$\bar{X}$	σ
.65 - 1.20	1.10	.131
1.30 - 1.80	.975	.23

Such a summary may be useful in smoothing the data or in reporting past behavior especially if there are reasons such as discussed in the earlier sections of the paper to expect a skewed pattern. Any prediction, however, should be fortified by some assurance that the data came from a controlled process.

In the present case there was reason to suspect that the skewness results from combining data representing several operating conditions. Using this hunch as a basis for selecting sub-groups the chart for averages shown in Figure 5 was prepared. In this chart the groups are plotted in order of decreasing sample size. Strictly speaking the control limits should be a series of steps instead of the sloping lines shown. The short cut shown is sufficiently precise to indicate significant differences among the sub-groups. The same conclusion could, of course, be reached by the more involved computations of the analysis of variance.

Thus ends the statistical study of the data. Following the clues suggested by this evidence, control charts using time as the basis of selecting the sub-groups have been installed with accompanying (and presumably consequent) improvement in uniformity.

5. INTERPRETATION

The interpretation in any specific instance is influenced by the nature of the reason for testing. In many cases proper interpretation may be aided by the statistical procedures, but leans heavily on familiarity with the technical field of study.

The example shows how the information given by each method supplements that from other sources. The third moment indicates a significant departure from normality which is depicted graphically by the plot on Hazen paper. More or less intuitive reasoning suggested the grouping of the data that resulted in the control chart. There is evidentally not one, but three levels of the data. At this point the study was concluded. It possibly could have been continued by further study as suggested by the disclosure of the three levels of the average, and may have lead to confirmation of the existence of three universes.

Restudy of the broad base that yielded the 130,000 observations has not been made. Perhaps this restudy never will be made, but the information even on this one time basis has been of value. An incidental by-product is this paper, for whatever it may be worth, bringing to your attention some of the reasoning and methods used in testing for normality.

CONCLUSION

There is an endless cycle of cause and effect among the factors involved in testing for normality. The reason for testing influences the selection of method. The method used influences the interpretation. The interpretation of previous studies is considered in planning for future testing.

Some uses of the third moment, of linear tests for skewness, of normal probability paper, and of control charts are illustrated for a distribution of selected production data.

Notes:

1. H.L. Reitz, Mathematical Statistics. The Open Court Publishing Co., Chicago (1927). Pg. 3.
2. John W. Dudley, Examination of Industrial Measurements, McGrap Hill (1946). Pg. 37.
3. The term essential information of a set of data is here used in the sense discussed in A.S.T.M. Manual on Quality Control of Materials.
4. A.L. Bowley Elements of Statistics, P.S. King & Son, London (1937).
5. Archie G. Worthing and Joseph Geffner, Treatment of Experimental Data, John Wiley and Sons (1943). Pg. 173.
6. NAVORD OSTD 80
   Naval Ordnance Standard Sampling Procedures and Tables for Inspection by Variables, Bureau of Ordnance, May 8, 1952, pg. 1.
7. V. Luzzati, Statistical Treatment of Errors in the Determination of Crystal Structures, Acta. Crystallographica 5, 802-10 (1952).

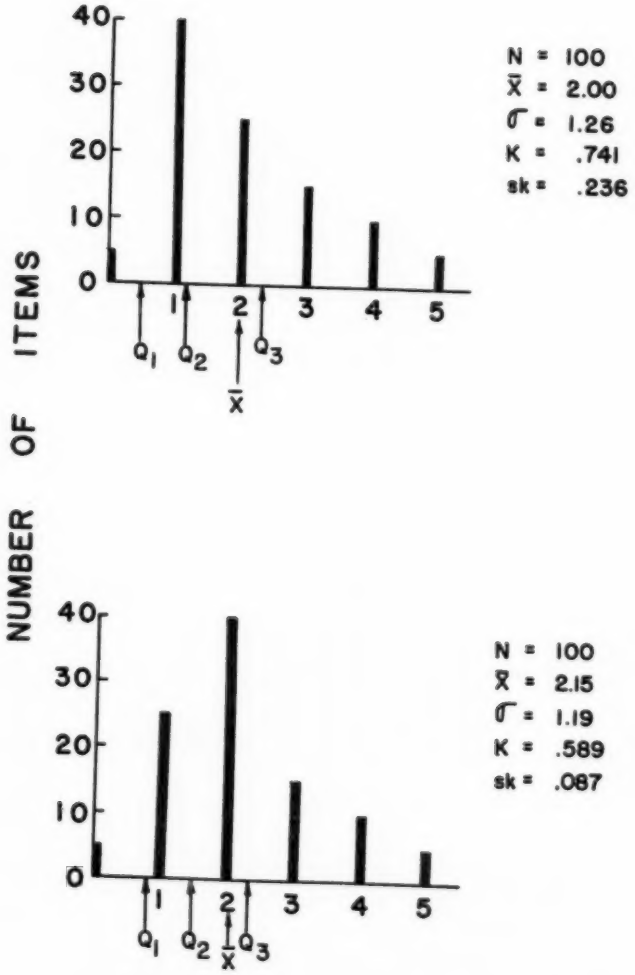


Fig. 1 Two groups selected to illustrate measures of skewness

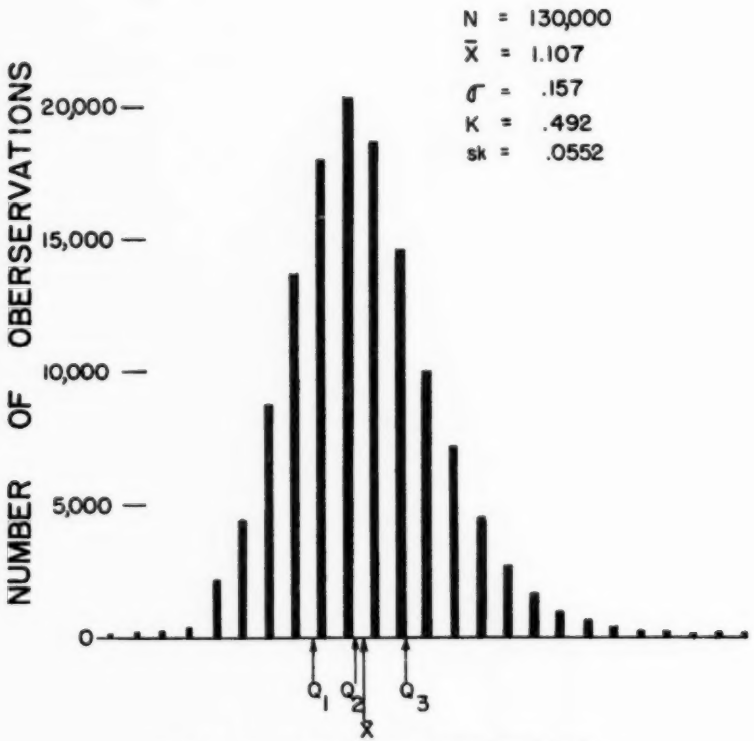


Fig. 2 Observed frequency distribution of a quality characteristic

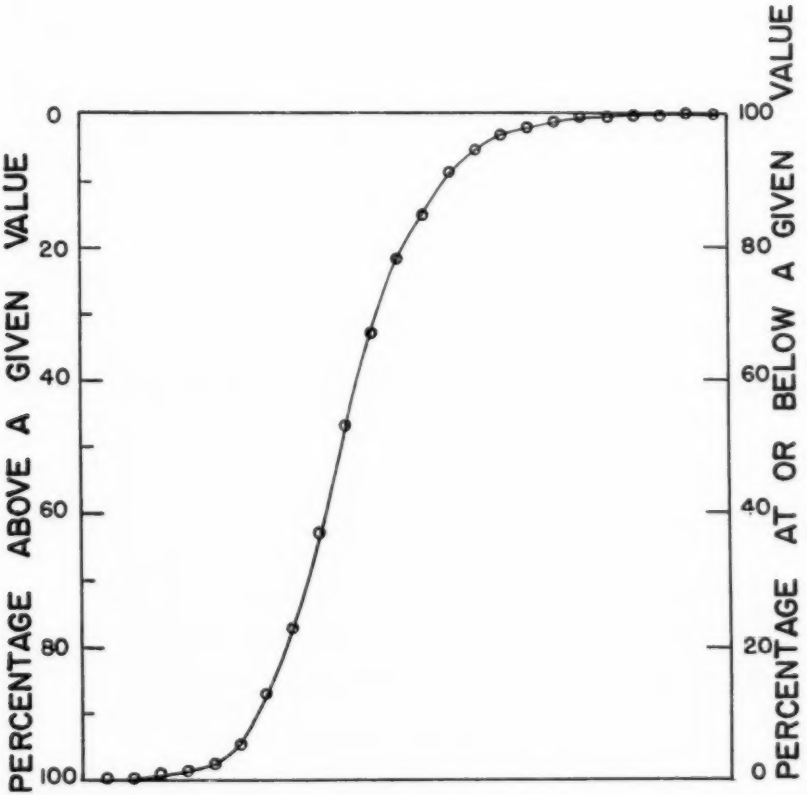


Fig. 3 Cumulative relative frequency distribution. data of fig. 2 plotted on rectangular scale

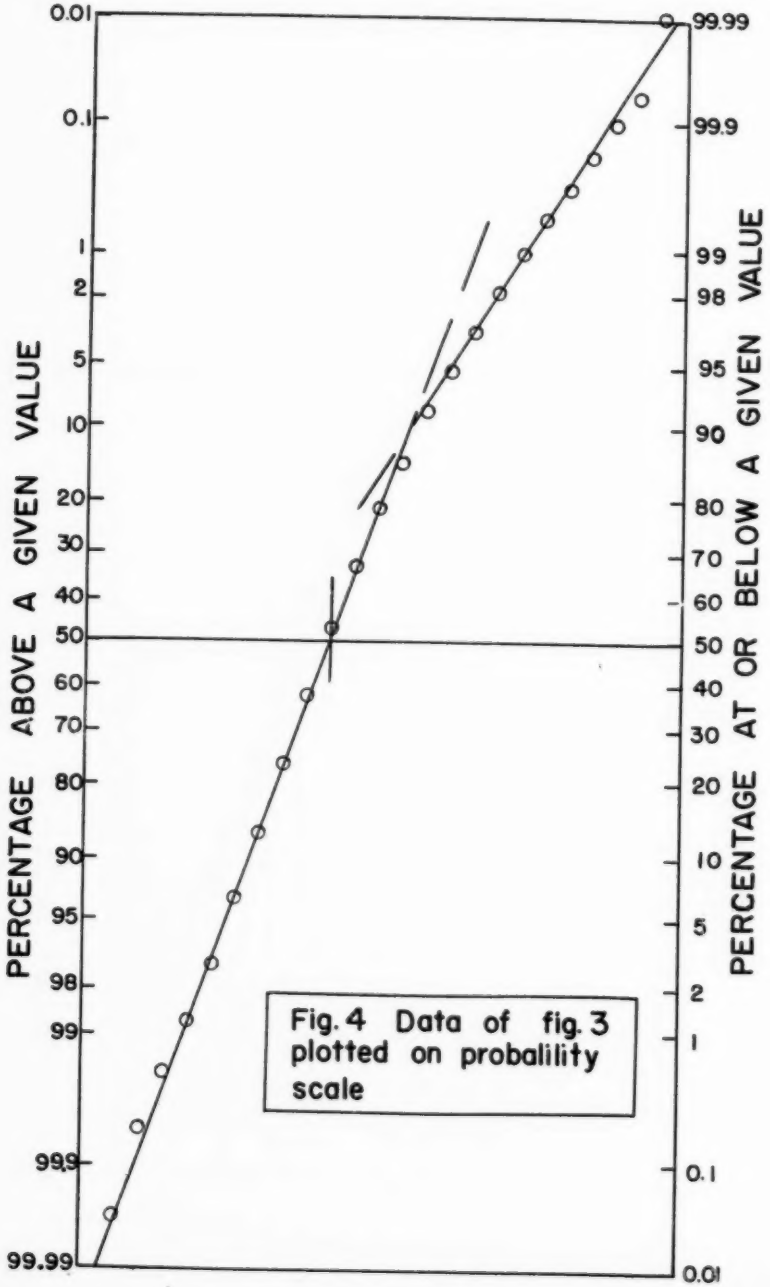


Fig. 4 Data of fig. 3 plotted on probability scale

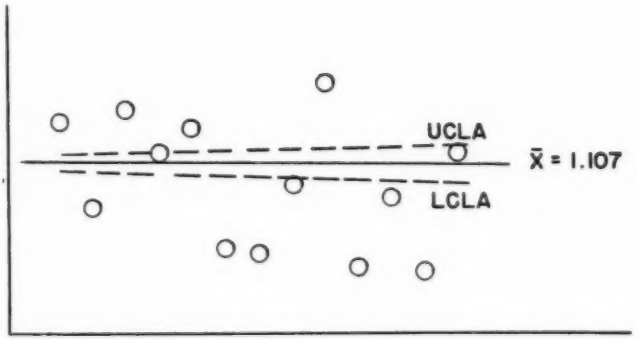


Fig. 5 Control chart for averages of rational sub groups from the distribution shown in fig. 2, 3, & 4

TABLE 1. LIMITING VALUES OF SKEWNESS

Sample Size n	Standard Error of k*	k	
		P = .01	P = .05
12	.637	1.64	1.250
24	.472	1.22	.925
48	.340	.875	.667
72	.281	.725	.551
96	.246	.634	.482
100	.242	.624	.474
500	.109	.281	.214
1000	.077	.199	.151

*The standard error of k is computed from the relation:

$$\sigma_k = 6n(n-1)(n-2)/(n-2)(n+1)(n+3)$$

Values of k less than the above limiting values might arise from sampling a symetrical population. Values of k greater than the limiting values support the inference that the population has skewness greater than zero.

TABLE 2. DATA FOR COMPUTING $\bar{x}$, σ and k
(See Reference 3, Table VIIb, pg.18)

Observed Values x 100	Cell Mid-point	Observed Frequency	Cumulative Frequencies: First Cum.	Second Cum.	Third Cum.
61/65	63	40	40	40	40
66/70	68	200	240	280	320
71/75	73	432	672	952	1272
76/80	78	736	1408	2360	3632
81/85	83	2102	3510	5870	9502
86/90	88	4432	7942	13812	23314
91/95	93	8719	16661	30473	53787
96/100	98	13619	30280	60753	114540
101/105	103	17903	48184	108936	223476
106/110	108	20280	68463	177399	400875
111/115	113	18677	87140	264539	665414
116/120	118	14501	101641	366180	1031594
121/125	123	9989	111630	477810	1509404
126/130	128	7119	118749	596559	2105963
131/135	133	4496	123245	719804	2825767
136/140	138	2700	125945	845749	3671516
141/145	143	1682	127627	973376	4644892
146/150	148	993	128620	1101996	5746888
151/155	153	612	129232	1231228	6978116
156/160	158	300	129532	1360760	8338876
161/165	163	222	129754	1490514	9829390
166/170	168	110	129864	1620378	11449768
171/175	173	65	129929	1750307	13200075
176/180	178	69	129998	1880305	15080380
181/185	B183	2	130000	2010305	17090685
Total		n 130000	2010305	17090685	104999486
Divided by n			F_1 15.46388	F_2 131.4668	F_3 807.688

TABLE 3. COMPUTATION OF AVERAGE AND σ DATA OF TABLE 2.

AVERAGE $\bar{x}$

B = Midpoint of highest cell = 1.83
$(F_1-1) \times .05 = 14.464 \times .05 =$ .72315
$\bar{x} =$ 1.10685
1.107 Rounded Value

STANDARD DEVIATION σ

$$\sigma = m\sqrt{2F_2 - F_1 - F_1^2}$$

$$m\sqrt{2F_2 - F_1(F_1 + 1)}$$

$$2F_2 = \frac{34,181,370}{130,000} = 262.9336$$

$F_1 = 15.46388$
$F_1+1 = 16.46388$

$15.46388 \times 16.46388 = 254.5954$
μ_2 8.3382 (Cell Units)
μ 2.8876 (Cell Units)

$\sigma = 2.8876 \times .05 = .14438$
$= .144$ Rounded

$\mu^3 = 2.8876^3 = 24.07739$

(For use in Table 4.)

TABLE 4. COMPUTATION OF SKEWNESS k.
Data of Table 2 and 3

$$k = - \frac{6(F_3-F_2) + F_1 - 3(2F_2-F_1)F_1 + 2F_1^3}{\mu^3 \text{ (Cell Units)}}$$

F_3 = 807.6880
F_2 = 131.4668
Difference 676.2212
Times 6 6
4057.3272
F_1 = 15.4639

F_1^3 = 3697.9882
F_1^3 Times 2 = 7395.976
11468.7671

$2F_2$ = 262.9336
F_1 = 15.46388
247.46972
Times 3
742.40916
Times 15.46639
- 11482.3896
+ 11468.7671
-13.6225

$$k = - \frac{-13.6225}{24.07739} = .565779$$

k = .566 Rounded

SAMPLING VARIATION OF AVERAGES, TESTING OF HYPOTHESES, AND SAMPLE SIZES

Norman R. Garner
Phila. QM Depot, General Testing Laboratories Division

I. Introduction

It is supposed that an experimenter takes a random sample of n individuals from a very large population of individuals, and that this population can be considered as being normally distributed or approximately normally distributed about some central value. Further, each individual is independent of the others. From this sample the experimenter can compute an estimated average ($\bar{X}$), an estimated standard deviation (s), or other statistics, such as the range (R).

On the basis of such data the experimenter usually desires to test the hypothesis that their sample could have originated from a specified or assumed population (called the null hypothesis H_0) against the hypothesis that this sample could not have originated from the specified or assumed population (called the alternative hypothesis H_1).

Or, the experimenter could have taken two such samples, assumed to have originated from the same population, and on the basis of these samples test whether or not this assumption is true. The following are some commom statistical tests employed to test such hypotheses.

II. A. Student's t-test (Population Mean Specified)

1. Two-Sided t-test

Suppose we draw samples of five from a population which is normally distributed. Let $\bar{X}$ be the computed sample average each time, and s the standard deviation. Then each time we compute the statistic $t = \frac{|\bar{X} - \bar{X}'|\sqrt{5}}{s}$, where $\bar{X}'$ is the assumed or specified population average. If a frequency chart of the t's were made we would have a distribution of the t values for samples of five. That is, we have a t-distribution or t-curve.

In the same manner in which we consider areas in under the normal curve and compute the chance of drawing any one individual greater than or less than a specified value, we consider the area in under the t-curve and compute the chance of getting a t value greater than or less than a specified $t = t_{\alpha}$. For example, (See Figure I) the chance of getting a t value greater than the absolute value of $t_{.05} = 2.77$ is 5%. Hence, we have a method for testing whether a sample could have originated from a specified population (H_0) or not (H_1). For example, a chemist weighed a sample of 10 Orthotolidine Tablets (water purification). His results were 304, 302, 304, 306, 309, 313, 309, 308, 291, 304. The calculated average weight per tablet is 305 mgs. The estimated standard deviation is 5.9 mgs. $\bar{X}'$ is supposed to be 300 mgs. So we test the null hypothesis (H_0) that the difference 305 - 300 is due to chance variation against the alternative hypothesis (H_1) that sampled average of 305 could not originate from the specified population centered about 300. We compute the statistic $t = \frac{|305 - 300|\sqrt{10}}{5.9} = 2.61$. Establishing the risk or level of significance at 5%, we find from the tabulated t values

for 9 d.f. that $t_{\alpha} = 2.262$, where $\alpha = .05$. (The number n with which to enter the table is determined by the number of degrees of freedom available for the estimate of s). Since $t > t_{\alpha}$, i.e. $2.61 > 2.26$ we say that there is sufficient evidence to reject the null hypothesis. Thus, it is concluded that a sample average of 305 mgs could not originate from a population of Orthotolidine Tablets for which the average weight per tablet is assumed to be 300 mgs and the estimated standard deviation is 5.9 mgs.

2. One-Sided t-test (Population Mean Specified)

On the other hand, it may be desirous to test the hypothesis that this sample could have originated from a population of Orthotolidine Tablets for which the population average is 300 mgs, against the alternative hypothesis that they originated from a population for which the average is greater (less) than 300 mgs. In this case we compute $t = \frac{(305 - 300)\sqrt{10}}{5.9} = +2.61$. For this one tailed test we have to use the tabulated t for 10% in order to test at the 5% level of significance since the tables are computed for the two tailed test (See Figure II). This value of $t_{\alpha} = 1.833$. Since $t > t_{\alpha}$, or $2.61 > 1.83$ we say that there is sufficient evidence to reject the null hypothesis. Thus, it is concluded that this sample originated from a population of tablets whose average is greater than 300 mgs.

B. Students t-test (Comparison of Two Sampled Averages)

1. Two-Sided Test

It is also known that

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$$

is distributed according to the t-distribution with $n_1 + n_2 - 2$ degrees of freedom. Hence we can test whether two samples of n_1 and n_2 individuals can be considered as originating from the same parent population (H_0), against the alternative hypothesis (H_1) that they originated from two different populations. In the above formula, $\bar{X}_1$, $\bar{X}_2$ and s are the estimated averages and standard deviations computed from the samples. For example, two technicians have sampled nine 4" x 6" swatches from a bolt of cloth, their results were

	Operator I	Operator II
	177	179
	181	162
	174	173
	177	168
	184	175
	187	172
	175	164
	193	177
	174	174
Sums	1622	1544
Averages	180.22	171.55
Variances	43.7	33.3

We wish to test the hypothesis that the two technicians are receiving equal breaking strengths against the alternative hypothesis that the

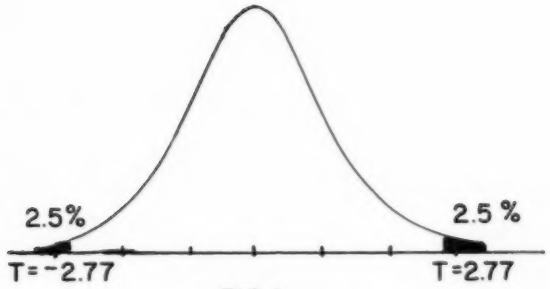


FIG. I

DISTRIBUTION OF T (N=5) TWO TAILED TEST

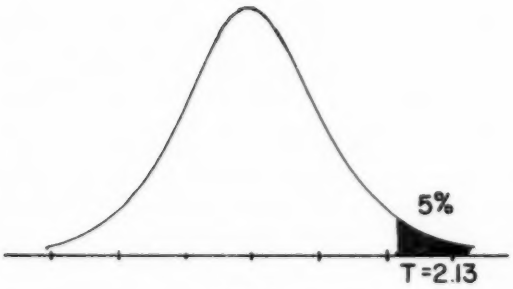


FIG.2

DISTRIBUTION OF T(N=5) ONE TAILED TEST

mean breaks of the two technicians are not equal. We compute

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|180.22 - 171.55|}{6.2\sqrt{\frac{2}{9}}} = 2.96$$

The tabular t for $n_1 + n_2 - 2 = 16$ degrees of freedom and a level of significance of 5% is 2.12. Since $t > t_\alpha$, i.e., $2.96 > 2.12$ we reject the hypothesis that the one technician is receiving breaks comparable to the other.

2. One-Sided Test (Comparison of Two Sampled Averages)

If we were testing the hypothesis that Operator I was equal to Operator II against the alternative hypothesis that Operator I is receiving breaks greater than Operator II, and using a level of significance of 5%, (again we consider the tabular t for 10%) we would have, for 16 d.f., $t_\alpha = 1.74$. Since $2.96 > 1.74$, we conclude that Operator I is getting breaking strength values greater than Operator II.

C. The Normal Test

If s_o is known, then the values of $t = \frac{\bar{X} - \bar{X}^1}{s_o}\sqrt{n}$ are distributed normally. Or if n is large, the t distribution approximates the normal distribution so closely that it can be considered normally distributed. (See Figure III) When this is the case, all that is needed is the normal deviates of the normal curve. This amounts to using the tabular values of the t distribution for infinite d.f. For example, if s_o for the weights of the Orthotolidine Tablets is 4 mgs, then, for the two-sided test, we have $t = \frac{305 - 300\sqrt{10}}{4} = 3.95$. The normal deviate for the 5% level of significance is 1.96. Since $t > t_\alpha$, that is, $3.95 > 1.96$ we reject the null hypothesis that the sample could have originated by chance variation from a population for which the true mean is 300 mgs. Also, for the one-sided test, the tabular value of the normal deviate is 1.64. Hence it is concluded that the sample originated from a population for which the true average weight per tablet is greater than 300 mgs. The procedure is the same for the comparison of two averages.

III. Remarks Concerning The t-test

It is well known that if values are normally distributed with variance s_o^2, then the mean of a random sample of n such items is normally distributed with variance s_o^2/n. Hence, even if the original distribution were not exactly normal, and since the means usually tend to normality as the size of the sample increases, application of the t-test can be applied widely and legitimately to cases in which we have not sufficient evidence to assert that the original distribution was normal, but in which case we have reason to think that it does not belong to the exceptional class of distributions for which the distribution of the mean does not tend to normality. If the original data is not normal and there is excessive skewness the two-tailed t-test is not too seriously jeopardized. When the experimenter thinks he is testing at the 5% level of significance he is probably testing anywheres from a 4% to a 7% level of significance. The 1% level probably would vary from ½% to 2%. However, as a rule, the level of significance is greater then 5% (1%). That is, too many significant results are obtained. For the one-tailed test excessive skewness may invalidate the use of the t-test. However, there

are ways of bringing into normality data which is not, simply by the use of transformations on the data before applying the t-test. One of the more useful transformations is the logarithmic transformation. That is, transform the variate x into another variate z by the transformation $z = \log x$.

IV. A Few Other Useful Tests

A. The Q And U Test

Schwartz, D.H. and Kaufman, Paul, "Determining Conformance With Requirements for Lot Average", Industrial Quality Control, Vol 7, No 5, March, 1951.

B. The use of the Range in Place of Standard Deviation in the t-test. Lord, E. "The Use of The Range in Place of Standard Deviation in The t-test". Biometrika, Vol 34, 1947.

C. The Use of Ranks to Avoid Assumptions of Normality. Freedman, Milton, "The Use of Ranks to Avoid The Assumption of Normality" Journal of The American Statistical Assn. Vol 32, 1937

V. Sample Size For Estimating The Average

Since $t = \frac{(\bar{X} - \bar{X}^1)\sqrt{n}}{s_o}$, we can compute the size of the sample required to estimate an average within certain limits for a prescribed level of significance. Suppose we wish to estimate the average weight per tablet of our lot of Orthotolidine Tablets within 2 mgs with a probability of .95. Then $1.96 = \frac{2\sqrt{n}}{4}$, so that $n = 16$. This procedure should be considered only as a quick estimate of n. More efficient means are given by Harris M., Horwitz, D.G., and Mood A.M. "On The Determination of Sample Sizes in Designing Experiments". Jour. Amer. Stat. Assoc. Vol 43, 1948.

VI. Type I and Type II Errors and Power of Test

If we test a hypothesis H_o by a test we may err by rejecting H_o when it is actually true. Such an error is called an error of the first kind (Type I). But we may also err by accepting H_o when it is false, that is when some other hypothesis H_1 is true. Such an error is called an error or the second kind (Type II). The probability of committing a Type II error is referred to as the operating characteristic of the test. The power of the test is defined to be the probability, or one might say the ability, that we reject H_o when it is false.

VII. Power Curves

A. One-sided Tests (Comparison of Two Averages)

We let the probability of the first kind, that is, of rejecting the null hypothesis when true, be P_1. This is usually called the "rejection level" or the "significance level". Let P_2 be the probability of making an error of the second kind. That is, of accepting the null hypothesis when it is false.

In testing hypothesis, a level of significance is assigned (say

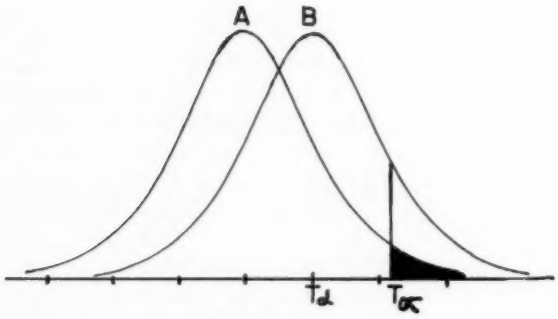


FIG. 5

DISTRIBUTION OF T UNDER H_0 AGAINST H_1 (ONE SIDED)

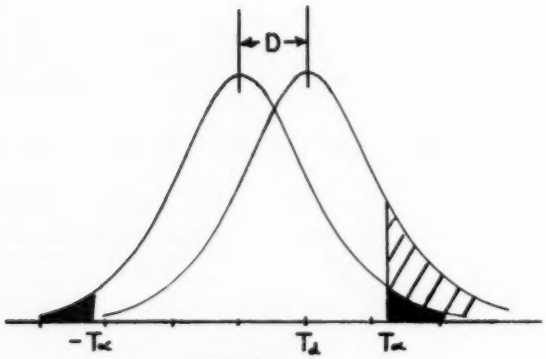


FIG. 6

DISTRIBUTION OF T UNDER H_0 AGAINST H_1 (TWO SIDED)

5%) and the null hypothesis is rejected if $t > t_{5\%}$ (or $t < -t_{5\%}$, depending on the direction of the test). For example, if the sample size for each of two samples was 16, and $t < 1.69$ we accept the null hypothesis. However, if D is the true difference, in standard units between the averages of the two samples, it is still possible to get a $t < 1.69$. Thus we would be making an error of the second kind, but what is the probability P_2. Figure IV gives a visual insight into the problem. The figure shows two curves: the t distribution under the null hypothesis, mean = 0 (curve A), and the t distribution when the true difference of our populations is D units (curve B). The shaded area to the right tail of curve A is the area of rejection. Whenever we get a value of $t < t_\alpha$ we accept the null hypothesis. Now if we have really sampled from two different populations, we shall be making an error of the second kind whenever we get a sample such that $t < t_\alpha$. The probability of this is approximately the area under curve B to the left of $t_\alpha - t_d$ and subtract from unity to get the area to the left of this point. That is, (t_d is the difference expressed in t units) the area in under curve A to the left of $t_\alpha - t_d$. Tables of t are usually given for fixed values of p so the exact value of P_2 may be difficult to find if the total degree of freedom $(n_1 + n_2 - 2)$ is less than 30. If $(n_1 + n_2 - 2) \geq 30$ normal curve tables may be used and interpolation will be unnecessary. However, P_2 may be pretty closely approximated, even if $(n_1 + n_2 - 2) < 30$, by interpolation in the t table. For example, if $\bar{X}_2$ is 1.5 t units above $\bar{X}_1$, $n_1 + n_2 - 2 = 18$, and $P_1 = .05$, t_α will be 1.734. We need to find the probability of a value of t greater than 1.734 - 1.50 = .234. From the same tables we find $t_{40} = .257$, and $t_{45} = .128$. Linear interpolation would give approximately $t_{43} = 239$, so $P_2 = 1 - .43 = 57$. This is the probability that if D = 1.5 t units the null hypothesis will be accepted when false. That is to say, that in a large number of experiments in which D = 1.5 t units only 43 per cent will lead to rejection of null hypothesis. Figure VI shows the Power Curve of the t test for this case. P_1 is .05 and $n_1 + n_2 - 2 = 18$. The exact values of P_2 were taken from the t table. By plotting $1 - P_2$ we are plotting the "power" of the test. P_2 itself is referred to as the operating characteristics. (Ferris, C.D., Grubbes, F.E., and Weaver, C.L., Operating Characteristics For The Common Statistical Tests of Significance. Annals of Mathematical Statistics, Vol 17, No 2, June 1946.

TABLE I

Data For Power Curve (One-sided Test)

$P = (1 - P_2)$	t	$t_d = (t_\alpha - t)$
.005	2.878	-1.154
.010	2.552	- .818
.050	1.734	- .000
.10	1.330	.404
.20	.862	.872
.30	.534	1.200
.40	.257	1.477
.50	.000	1.734
.60	- .257	1.991
.70	- .534	2.268
.80	- .862	2.596
.90	-1.330	3.064
.95	-1.734	3.468
.99	-2.552	4.286
.995	-2.878	4.612

It is necessary to point out that this procedure for computing the power of the test is approximate, however this approximation is very good. Exact procedures become rather difficult and the literature on this subject should be investigated. Nevertheless, from a practical utilization the above procedure is adequate. The main flaw in the argument is the utilization of s instead of s_o. This approximate method depends vitally upon a very good estimate of s_o.

B. Two-Sided Test

In this test the null hypothesis is rejected when the absolute value of t is large. (See Figure V). In our case, with 18 degrees of freedom and $P_1 = .05$ we reject the null hypothesis whenever we get a value of $|t| > 2.101$. The probability of rejecting the null hypothesis when the true average is D t units greater than or less than the t for the null hypothesis is equal to the area to the left of $-t_\alpha$ and to the right t_α of curve B. To compute our power curve we first compute P', equal to the area to the right of t_α, then P", the area to the left of $-t_\alpha$. ($t_\alpha = 2.101$). We do this by setting the area in under the curve to the right for a given P'. For this area compute what the difference is in t units. Then for this difference, find the area to the left of $-t_\alpha$. Adding P' + P" gives us the power of the test. Figure VI shows the power curve for this test.

TABLE II

Data For Power Curve (Two-Sided Test)

$P = (1 - P_2)$	$t_1 = t_d - t_\alpha$	t_d	$t_2 = t_d + t_\alpha$	P"	P' + P" = P
.025	2.101	0	2.101	.025	.05
.05	1.734	.367	2.468	.010	.06
.10	1.330	.771	2.872	.005	.105
.20	.862	1.239	3.340	.005	.20
.30	.534	1.567			.30
.40	.257	1.844			.40
.50	0.000	2.101			.50
.60	- .257	2.358			.60
.70	- .534	2.634			.70
.80	- .862	2.963	Negligible		.80
.90	-1.330	3.431			.90
.95	-1.734	3.835			.95
.975	-2.101	4.202			.975

VIII. Sample Size

The question is now raised as to the appropriate sample size. It should be noticed that by considering P_1, P_2, D, n, as soon as three of these are given, the fourth is fixed. Usually P_1 and D are assigned. Hence the answer as to sample size can be solved either by fixing P_2 and solving for n, or fixing n and solving for P_2. Then, if P_2 is not suitable, assign another value to n and solve again for P_2. Keep doing this until you are satisfied with the value of P_2. Remember, that all methods depend vitally upon some reliable estimate of the standard deviation. Since, when testing a symmetrial two sided hypotheses, and if P_2 is not too high, the area to the left of $-t_\alpha$ becomes negligible. Hence very good approximations can be made. We first consider the one-sided tests. Consider the Orthotolidine Tablets. From past experience, a good estimate of s is 4 mgs. It is desirous to detect a difference of 5 mgs between two lots. $P_1 = .05$. If we take samples of 10 for each lot, then we have 18 d.f. for the estimate of s. We solve for t from the equation.

$$d = \sqrt{\frac{2}{n}}\ s\ (t_\alpha + t)$$

$$5 = \sqrt{\frac{2}{10}}\ (4)\ \ (1.734 + t)$$

$$t = 1.066$$

Therefore the probability of catching the difference is $P = 1 - \frac{1}{2}(30) = .85$. That is, $P_2 = .15$, and the power of the test is .85. If this is satisfactory, we stop. If it isn't, increase n.

Next, we can utilize tables prepared by Neyman, S. and Tokarska, B. "Errors of The Second Kind in Testing 'Students' Hypothesis". Journal of The American Statistical Association, Vol 31, 1936. These tables give the standardized error (they use the symbol ρ) for given P_1, P_2, and n. The differences which it is desired to detect and the standard deviation are expressed in terms of percentage of the general mean. In the case of the Orthotolidine Tablets, $s_o = 1.33\%$ of the general mean so that the standard error of the difference is $s_d = .596$ of the general mean, and $D = 1.66\%$ of the general mean. Then, for these tables, $D = \rho\ s$, so that $\rho = 2.78$. If $P_1 = .05$, $N = 18$, the chance of non-detection is .16, so that the power of the test is .84. We see that either method gives good results.

We now consider the sample size for the two-sided test. If we are interested in the two-sided test for our two lots of Orthotolidine Tablets, then the difference which it is desired to detect expressed in standard units is $t_d = \frac{5}{4}\sqrt{5} = 2.80$. Again, we let the level of significance be 5%, so that $t_\alpha = 2.101$. Then $t_1 = t_d - t_\alpha = .699$. From the tabulated values of t we find that $P_2 = \frac{1}{2}(.48) = .24$ and hence $P' = .76$. Then $t_1 = t_d + t_\alpha = 4.901$ so that P'' is negligible. Therefore, the power of the test is $P' + P'' = .76 + 0 = .76$. It is noted that the one-sided test is more powerful than the two-sided test. Again we state that if .76 is sufficient, no more calculations are necessary. If a greater power is desired, increase n. Further, we can utilize tables prepared by Ferris, C.B., Grubbs, F.E., and Weaver C.L., "Operating Characteristics For The Common Statistical Tests of Significance". The Annals of Mathematical Statistics Vol 17, No 2, June 1946 p 195. They have determined the operating characteristics in terms of

$$\lambda = \frac{\frac{D}{s}}{\sqrt{\frac{n}{\frac{n_1 n_2}{n_1 + n_2}}}}$$

and n, where $n = n_1 + n_2 - 1$. For our case, $\lambda = .642$ and hence P_2 read from their chart, is approximately .25, so that the power of the test is .75. Again we have good agreement.

IX. Further Selected Bibliography

Fisher, R.A., and Yates, F., "Statistical Tables", London: Oliver and Boyd, 1938. Bartlett, M.S. "The Use of Transformation". Biometrics 3, pp 39 - 52, 1947. Walsh, J.E. "On The Information Lost By Using A T-Test When The Population Variance is Known" Jour. Amer. Stat. Assoc. 44, pp 122 - 125, 1949.

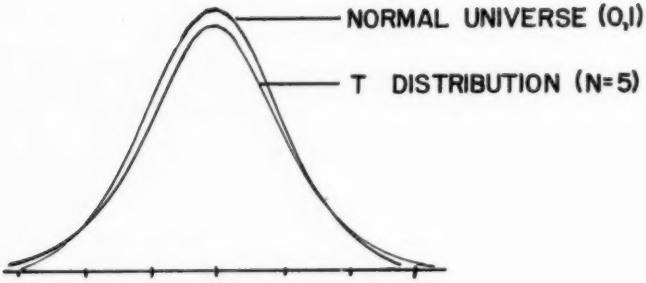


FIG.3

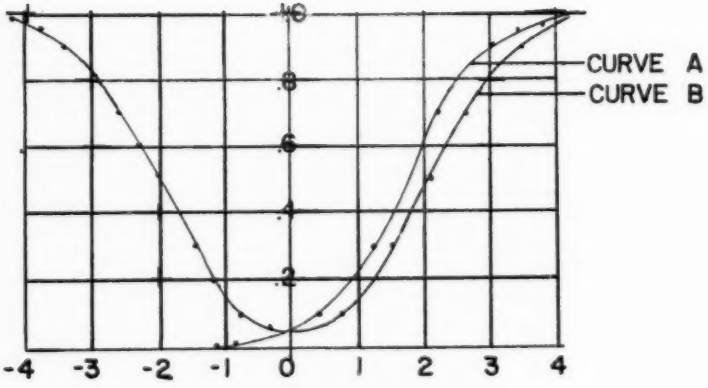


FIG.4
TRUE DIFFERENCE IN STANDARD UNITS.
POWER CURVE OF T-TEST. DEGREES OF FREEDOM ($N_1+N_2-2=18$). $P_1=.05$.
CURVE A IS ONE SIDED TEST. CURVE B IS TWO SIDED.

WHO CONTROLS THE VENDORS' QUALITY

J. Philip Worth
Electrolux Corporation

Will you please take a trip with me down memory lane? I can remember with a rather nostalgic feeling, a street corner of a bustling city. Standing on this corner in the same spot for many years has been a small two wheeled cart. This cart is as non-descript as the attendant, but the whistle of the cart sings merrily. This is the attraction and the advertising of the peanut vendor. Such advertising is tempting, as so are the smiles of the vendor and the customers as they walk away with their small bag of fresh, warm, tasty, peanuts.

Today the peanut vendor is gone. An investigation of his disappearance has disclosed:

1. Some customers complained about the varying taste quality of the nuts, so the vendor revealed he had roasting troubles, but it would be corrected. - Nothing was done.
2. When taste quality complaints persisted, the vendor was then having trouble with his supplier on the grade of peanuts. - Still nothing was done.
3. As the complaints mounted, so did the excuses, sales dropped off, and business went from bad to worse until it disappeared.

I know you do not sell peanuts on a street corner, but you do sell a product of which you are proud and to which you hope to be associated with for some time. Therefore, I have tried to develop this picture, because it shows the fundamental influence exerted by the customer over your product and also shows why you should control the Vendors' quality.

Modern manufacturing techniques have long since passed the master craftsman stage in their quest for mass production. Each phase of the master craftsman technique has been specialized so that mass production can be fulfilled by teamwork between the sales engineer, design engineer, manufacturing engineer, and the quality control engineer. By implication it necessarily follows that to this group falls the full management task of producing and delivering to the customer what he desires. That part of such a management task about which we are particularly interested is determined by elimination. The sales, design, and manufacturing engineers are busy selling and manufacturing while the quality group are required to ascertain and assure delivery of only what the customer desires. Therefore, you, the quality personnel, become the customers' representative throughout the whole process.

With these thoughts in mind, I propose to discuss:

1. Customer influence throughout in-plant activities.

For instance, armed with engineering specifications and data, final inspection is required to supply the final test plan for customer quality assurance to the product before shipment. As quality fluctuations appear, you, the quality group must extend your customer influence to final assembly with its own operations as well as the parts delivered for assembly.

In many cases, quality follow-up is necessary from final inspection back through assembly, fabrication, and receiving inspection, until you knock on the purchasing department door.

2. Customer influence at receiving inspection and purchasing departments.

When quality fluctuations have been traced to variations in a vendors product, as defined by the receiving inspection data, you are already in the first stage of controlling the vendors quality. For instance, you should be certain that the purchasing department has recorded your findings with the vendor, and not as the peanut vendor failed to do. If you continue to inform the purchasing agent of these variations, and back them up with data, you will overcome the natural reactions, - the so-called "purchasing perogatives". As your program develops, you will notice a gradual change of attitude and will soon find that you will become an important part of the purchasing function. This development will become obvious when the purchasing agent asks for your data or information or opinions in his contacts with the vendor.

At this point the purchasing agent has recognized that you:

a. Know the specification of the material.
b. Have provided the inspection plan and testing equipment.
c. Are familiar with the fabricating operations and customers demands.
d. Receive first hand detailed information on all materials received.
e. Can furnish important inspection data.
f. Are able to provide liaison between customer and vendor for quality control purposes.
g. Can establish an effective control at the Vendors' plant.

This is the completion of the second stage of controlling the Vendors' quality while still representing the customer.

3. Customer influence at the Vendor.

Self interest by the purchasing agent will dictate that you should aid his cause of procurement. This will no doubt come when he is obtaining deliveries of inferior grade materials, spasmodic deliveries, or no deliveries at all. At this point you will be asked to visit the vendors facility and again be the customers representative. This time, however, the customer is the purchasing agent.

You now can put into action the executive skill for which you have been trained. As this is the third stage of controlling the vendors quality, you are ideally situated where you can improve your own position and be an important link between consumer and vendor. This should be done by a carefully planned program of consumer-vendor liaison for the promotion of mutual progress and responsibility. As such a program unfolds:

a. The vendors' quality will improve.
b. The vendors' deliveries will improve.
c. Fabrication and assembly operations will improve.
d. Final inspection results will improve.
e. Outgoing quality and the customers reaction will improve.

4. A case history showing customer influence in five natural steps.

After several successful years of business, a well-known motor manufacturer realized that its existing manufacturing plan would not produce the maximum quality and quantity required by its sales department, except at a substantial increase in unit cost. One of the outstanding examples is that of receiving inspection, machining, assembling, and finishing of commutators for the motor armature.

The existing plan provided for commutators to be assembled by the vendor, tested by receiving inspection, and processed in the factory. When quality fluctuations appeared with expanding production, improved operational sequences, and higher labor rates, faster action was required for the reduction or elimination of quality variances.

First Step - Customer Influence (quality dept.) Enters Vendors' Plant.

The quality department being familiar with the type, variety, and magnitude of variance, was requested to take action for faster permanent correction. The entry of the quality department was made through the vendors top management with permission to work directly with manufacturing personnel from the Plant Manager down.

Second Step - Consumer-Vendor Quality Agreement.

During the first meeting at the Vendors' plant, the conditions of quality variance were discussed in detail to show why corrections were required. In addition, the quality department (customer influence) outlined a plan for liaison by installation of an inspector who would be in first hand contact with all phases of the quality function in both plants. These were accepted as fact, but without too much enthusiasm because of past history and normal resistance to change.

Although the idea of the liaison plan was accepted, much effort had to be applied on improved quality standards. These were finally agreed upon, vendors' inspectors trained, and the quality level of shipments from the vendor immediately started to improve.

At this point the customer influence through you has extended to the fourth phase of controlling the vendors quality.

Third Step - Consumer-Vendor Shop Orientation.

Measured by former standards of management it would appear that we have extended the customer influence as far as possible. Today that is not so. With this in mind, the vendors top management was contacted to discuss the complete orientation of his key personnel and for a program to eliminate rejections in his plant. Upon agreement, the Plant Manager and all key personnel visited the consumers plant to acquire full knowledge of the consumers' processes, the products' end use and customers' requisites. This is the extension of customer influence to the fifth stage of controlling the vendors' quality.

Fourth Step - Consumer-Vendor Preventative Inspection.

Customer influence exemplified by the consumers gages and test equipment impressed the vendors' personnel sufficiently so that he provided gage and test equipment which paralleled that of the customer.

As variations between test set-ups disappeared the Vendor realized that quality fluctuations were apparent either in his process or his purchased material. At this point the liaison man was contacted for suggestions and mutual action.

The results of these conferences were:

a. Analysis of quality fluctuations.
b. Establishment of an improved process inspection.
c. Establishment of an improved receiving inspection.
d. Establishment of a periodic tool inspection.
e. Purchase of standby tools which were duplicates of existing ones.

At this point, customer influence has extended to its sixth stage.

Fifth Step - Consumer-Vendor-Vendor Relations.

When it became apparent to the Vendor that his processes could not remain in control due to variations in his purchased parts, he again asked for help from the liaison man. Such help was immediately forthcoming in the form of visits with the vendor to his vendor. Some of the results of these visits were:

a. Improved tools - From tool steel to carbide.
b. Additional tools to replace others as they became worn.
c. Larger production runs to eliminate set-up variances.
d. Changes in vendors' designs to improve safety and prevent test losses.

In this case history, you can see that the customers influence has extended through six stages from the time of your receiving inspection reports and on through five steps of controlling the vendors quality back to his suppliers. It should be stated here, that as a result of these activities, over 90% of the consumers receiving inspection costs on these commutators have been eliminated. In addition, the vendors' quality is sufficiently controlled to allow for the establishment of a unique sampling plan. This plan provides for a sample selection from each shipment by the liaison man who delivers the sample to the consumer for testing. If the sample meets all requirements, the liaison man issues instructions for the release of the shipment. If the sample should be unsatisfactory, the shipment is retained by the vendor for further treatment and submission of new samples.

5. A suggestion that you control your vendors' quality.

When you leave this convention and return to your own problems, analize them to see where you can best help your purchasing agent. He has many problems procuring the right material at the right time and at the right price. You can give him all the right details concerning quality. If you do it in the right way it will be allright with the purchasing agent for you to make the vendor quality right the first time.

SOME APPLICATIONS OF STATISTICS IN A REFINERY

C. T. Shewell
Humble Oil & Refining Company

The Humble Oil & Refining Company became interested in the application of statistical techniques to refining problems about three years ago. Since that time it has afforded opportunities to about twenty people to gain varying degrees of proficiency in statistics by taking advantage of the various short courses offered by such universities as Illinois, Purdue and Columbia and by bringing to Baytown such recognized authorities as Professors Tukey and Ostle to conduct intensive training courses. These courses have not produced any expert applied or theoretical statisticians, but they have opened our eyes to the many possibilities where statistics can help solve refining problems, both in the plant and in associated research work.

Some of our earliest attempts to apply statistics to refinery problems were in the determination of the standard deviations of various test methods, such as octane number, vapor pressure, viscosity, flash, distillation and so on. We were not interested in what the best tester could do when forewarned that a particular sample was a check sample, but we were interested in what went on every day in the routine operation of the laboratory, since this is the basis on which our business is run.

As a typical example of this phase of our work, let us consider in detail the program for determining the standard deviation of the Pensky-Martens flash point of an oil in the high temperature range. The laboratory foreman normally had six people who ran high P-M flashes, these people covering 24 hours a day, seven days a week, so that if we obtained data on all of them we would have complete coverage on this test. It was decided to use four oils which fairly well covered the flash range, and to get three tests by each tester on each oil over a period of several weeks in order to include the effect of what time of day the tests were run, but not so this factor could be evaluated separately. We would thus obtain data as to whether the testers all tested alike and whether the flash level of the oil influenced the standard deviation of testing. The results of this program are shown in Table 1, where it is pretty evident that tester #5 was poor on repeating himself and tester #1 wasn't too much better, and tester #5 tended to be below the average and #1 above the average. Tester #6 was able to reproduce himself much better than any of the others.

These results were so bad that a thorough investigation was launched of the practices in the flash room, leading to some changes and also to a retraining of all the testers. About six months later a similar program was undertaken, the data being shown in Table 2. It can be seen at once that quite a respectable improvement has been realized, the standard deviations being cut about in half. Table 3 shows the analysis of variance of these data on the assumption of a pure Model I.

It will be seen that with the interaction of oils and testers being significant at the 1% level, additional training was necessary, as finding significance here means that a tester was not always high, for example, but was low on high flash samples or high on low flash samples, whereas another tester was the reverse. Had a mixed model or pure Model

II been assumed, the variance between testers would not have been significant, but it appears here to be significant at the 5% level; this is additional evidence that more training was needed.

At about the same time that the testing work was being done we became interested in improving the operation of a fractionating tower producing a material of not less than 85% purity which we sell to a chemical company. The practice was at that time to take a sample of the product every two hours, analyze it, and adjust the tower controls as indicated by the analysis. Whenever the analysis showed less than 85% purity the product was sent to slop tankage and continued to be sent there until a subsequent analysis, two or more hours later, showed 85% or better purity, when the product would again be sent to finished storage. The familiar $\bar{X}$-R Chart for a typical run is shown in Figure 1. Quite obviously the process was not in control, and since no premium was paid for quality in excess of 85%, too much loss in yield was being suffered through the high product purity. The actual purity in finished storage was higher than shown, since about 15% of the product was sent to slop during periods of low indicated purity.

It was suspected that part of the large range was caused by the relatively poor accuracy of the analytical method, so a better method was worked out by our research group, which reduced this source of error to about one half its former value.

Meetings were then held with the operating staff to explain the principles of quality control and agreement was reached that so long as the quality of the finished product was above 85% there was no valid reason for cutting any production to slop.

Figure 2 shows the operation as of July, 1951; it will be noted that only one point on the $\bar{X}$ Chart was out of control of much narrower control limits. Although not shown on the chart, there was a period of 108 hours in which no adjustments of any kind were made to any of the tower controls, and no slop at all was produced, even though the average quality was very close to the specification.

Every one who owns a car buys lube oil in cans, and it is up to the refiner to see that you get the amount you pay for. A can filling machine works on a volumetric basis, but in order for the oil to flow readily into the cans it is placed in the machine reservoir at an elevated temperature, and a quart of oil measured at 150°F. is not a quart of oil as defined by State bureaus of weights and measures. As a matter of fact, some states require that a can of oil, when drained for one minute at a temperature of 60°F. shall deliver one standard quart of oil into your crankcase.

Since lube oil is an expensive product, it is incumbent on the refiner not to overfill the cans, and the law says he cannot underfill so he has two pinch points within which to operate his canning plant. A complication is the necessary allowance for drainage, which varies with the grade of oil, the heavier oils requiring a larger allowance for draining. The empty cans are another source of trouble because neither they nor their lids have a constant weight, and because the cans are filled hot, a check on their proper filling must be made by weight.

Our can filling machine was studied as it had been operating with the results shown in Figure 3. There are 24 filling pockets, each filling one can per cycle, and it is very evident that while each pocket was very reproducible in its own performance, the different pockets did not perform alike. Note also that the empty cans varied about 10 grams in weight, and the maximum pocket variation was 11 grams. For this oil the necessary holdup allowance was 16 grams.

The machine operation was so poor (about 1-1/2% overfill necessary on the high fill pockets to be assured of legal fill on the low fill pockets) that the manufacturer was called in to adjust it before any further work could be done. Subsequent to this adjustment another series of cans was examined with the results shown in Figure 4. There is still some evidence of differences between pockets but the range is only six grams, so that maximum overfill is now about 3/4%.

From these, and additional data, the standard deviations of empty can weights, and within pockets and between pockets oil weights were calculated, and using the formula for propagation of error, the combined standard deviation for these three sources of variation was found to be 2.15 grams. The test scales at the filling machine now have a tare weight equal to the sum of the weights of (1) an average can, (2) necessary drainage, (3) one standard quart of oil, (4) three standard deviations, or 6.45 grams. Over- or underfill is shown by a random can weighing 6.5 grams more or less than the tare weight.

In producing a product such as motor gasoline the refinery will process three to a dozen different base stocks which are finally blended together in proper proportions to give the customer a balanced motor fuel with the desired ease of starting, quick warm up, freedom from knock and vapor lock, and economy of operation that the modern motorist demands. This blending operation is not one to be undertaken by a novice as the desired qualities of the base stocks do not blend in simple arithmetical ratio to their volumes in the final blend.

Even when the blender makes a blend that exactly meets the specification, testing errors cause such blends to be rejected half the time as not meeting the specification, and many blends not meeting the specification will be accepted by test. When a blend is rejected by test it is necessary to alter its composition by adding the appropriate amount of one or more of the base stocks and resubmit it for test. This reblending costs money, as does the testing, and if too high a quality is made in an effort to save the reblending costs, then this excess quality cost becomes large. Somewhere above specification quality lies the point at which these two cost sources become equal, and this point should be the blender's target, at which he aims for the quality of every blend. Even though he aims at target quality, the blender will actually have a more or less normal distribution of blend qualities about the target. Repeated testing (10 or more per blend) will give an average whose value approaches that of the true quality of each blend. When some 35 or more blends are so tested, a measure of the blender's standard deviation (σ_B, may be obtained, and from the replicate tests there will be a measure of the testing standard deviation (σ_T). With a constant σ_B, σ_T, cost of reblending and cost of excess product quality (the cost of target quality compared to specification quality) there is a simple way to reduce the total cost of the product - reduce the excess quality and increase the

number of tests per blend. The complex relationships that link these factors together are shown in Figures 5 and 6 calculated on IBM's-CPC assuming normal distributions, where manufacturing tolerance is twice the distance from specification to target, because very high quality blends should also be reblended. Per cent bad product is that proportion of all blends accepted by test which would be shown to be actually below specification quality on exhaustive testing by current methods, or once by some hypothetical perfect test method.

Having decided what per cent bad product can be tolerated, a series of calculations are made based on the costs of N (the number of tests), manufacturing tolerance, and reblends, using of course the values of σ_B and σ_T that apply in your own process. These costs will in general be found to lie on curves that have a minimum cost at fairly high values of N, and it will usually be found that some place at lower values of N than the optimum will be most expedient from a practical standpoint. Savings of $200 to $500 per day are not uncommon using this method.

Another field in which we have applied statistics in the refinery is in the design of experiments. As an example, we are working closely with the American Petroleum Institute in an effort to design an air flotation separator which will remove oil and suspended particles of various chemical composition from the waste water of a refinery.

Humble has constructed a pilot air flotation unit which takes suction on our final separator and has provisions for varying the amount of air, water, settling time, type and amount of floc, pH of the treated water, and time of contact of air and water before discharge to the settling basin.

It appears that there are seven variables to investigate, and if only two levels of each variable were examined a completely randomized factorial arrangement would require 2^7 or 128 runs, and three levels would take 3^7 or 2187 runs. When the magnitude of these numbers was realized, efforts were made to find out whether some of these variables had no effect in what was anticipated to be the normal operating range. It was soon found that air in excess of that required to saturate the water was of no use and that this could be assured at a given time of contact in the saturating drum, and as water rate was the control over this, it was set constant. It was decided to study only alum at this time as a floc, thus leaving only settling time, pH, and floc concentration as variables.

With this reduction in the number of variables it became possible to consider a larger number of levels of each variable, and the final design was a 4x4x3 completely randomized factorial experiment. The major reason for the randomization was the known variability of the quality of the water entering the pilot unit; that is, its oil and suspended solids content was an uncontrolled variable that would certainly affect the performance of the pilot unit even under duplicate treating conditions.

Because of the variation in quality of the charge water we were unable to analyze the data ourselves, and called in Professor Tukey to help us interpret it. His interpretation made it clear that additional work was needed to test the effect of other flocs at different pH levels, whereas the settling time appeared to be unimportant above a certain value, this being caused by the high rate of rise of the air bubbles.

Another application of the principles of experimental design has been in cooperation with the Coordinating Research Council desert storage program for stability of gasoline. Two types of five gallon blitz cans are in the study, with four cans of each type filled with every mixture, two antioxidants each at two concentration levels, 10 fuels, four sampling periods over a two year span, and eight cooperating testing laboratories.

Were each laboratory to test every sample the work load would be tremendous (about 4000 tests per laboratory) so a balanced incomplete block design was worked out which reduced the work load per laboratory to about 400 tests. A similar program has been set up for Jet engine fuels, but on a smaller scale.

The foregoing brief descriptions of some of the work in the field of statistics currently being carried on at Humble Oil & Refining Company's Baytown refinery may have given you some idea of the possible scope of the varied applications of these techniques to the solution of refining problems.

Table 1

Test Results of High P-M Flash Study

Oil

Tester	A	B	C	D	Totals
1	505	460	415	485	
	535	455	365	465	
	500	480	380	495	
	1540	1395	1160	1445	5540
2	505	460	370	485	
	490	450	350	440	
	495	450	360	475	
	1490	1360	1080	1400	5330
3	490	430	360	460	
	520	440	365	490	
	520	460	365	495	
	1530	1330	1090	1445	5395
4	525	445	360	485	
	520	440	360	490	
	530	470	365	485	
	1575	1355	1085	1460	5475
5	505	465	365	440	
	425	450	365	490	
	435	460	380	490	
	1365	1375	1110	1420	5270
6	530	455	360	490	
	525	460	365	485	
	525	460	565	485	
	1580	1375	1090	1460	5505
Totals	9080	8190	6615	8630	32515
Average	504.4	455.0	367.5	479.4	
s	21.0	11.3	12.0	18.2	

Table 2

Test Results of High P-M Flash Study
After Retraining

Oil

Tester	E	F	G	H	Totals
	390	490	450	545	
1	400	480	445	525	
	380	480	435	505	
	1170	1450	1330	1575	5525
	380	490	460	520	
2	380	480	450	530	
	385	480	445	530	
	1145	1450	1355	1580	5530
	390	475	455	515	
3	375	480	455	485	
	395	480	455	485	
	1160	1435	1365	1485	5445
	380	485	450	520	
4	375	475	455	525	
	380	470	450	525	
	1135	1430	1355	1570	5490
	400	470	450	530	
5	390	480	460	515	
	390	475	455	530	
	1180	1425	1365	1575	5545
	385	485	450	535	
6	410	485	450	530	
	385	490	455	535	
	1180	1460	1355	1600	5595
Totals	6970	8650	8125	9385	33130
Average	387.2	480.6	451.4	521.4	
s	8.8	5.3	5.1	11.7	

Table 3

Analysis of Variance of Data in Table 2
Assuming Model I

Source	S.S.	d.f.	M.S.	F
Oils	172112.50	3	57370.83	
Testers	1073.62	5	214.72	3.19*
O x T	2779.16	15	185.28	2.75**
Error	3233.34	48	67.36	
Total	179198.62	71		

* Significant at the 5 per cent level.

** Significant at the 1 per cent level.

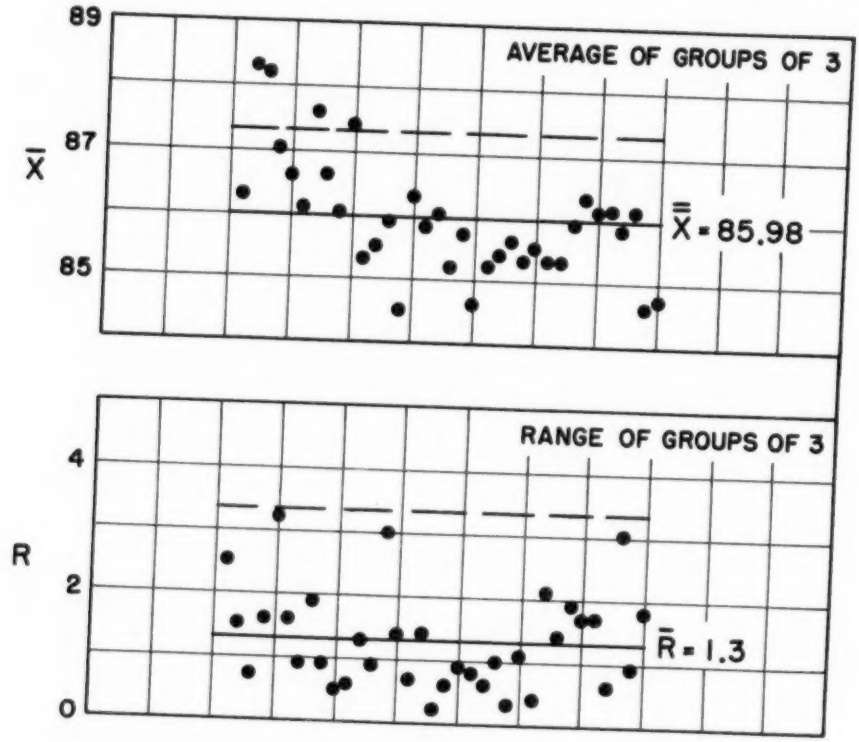


Fig. 1. Original Production.

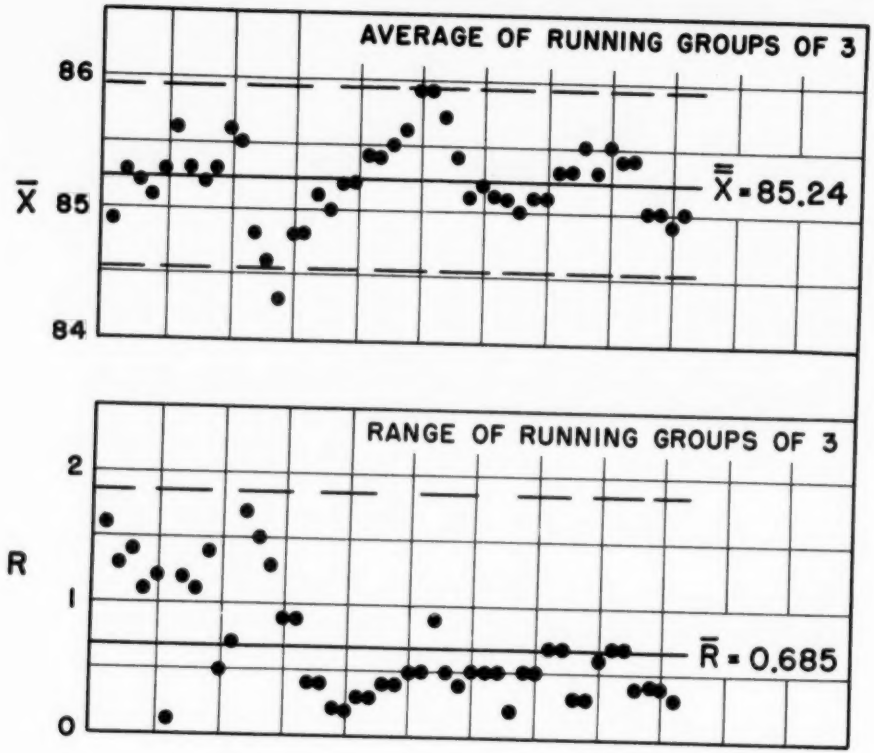


Fig. 2. Controlled Production.

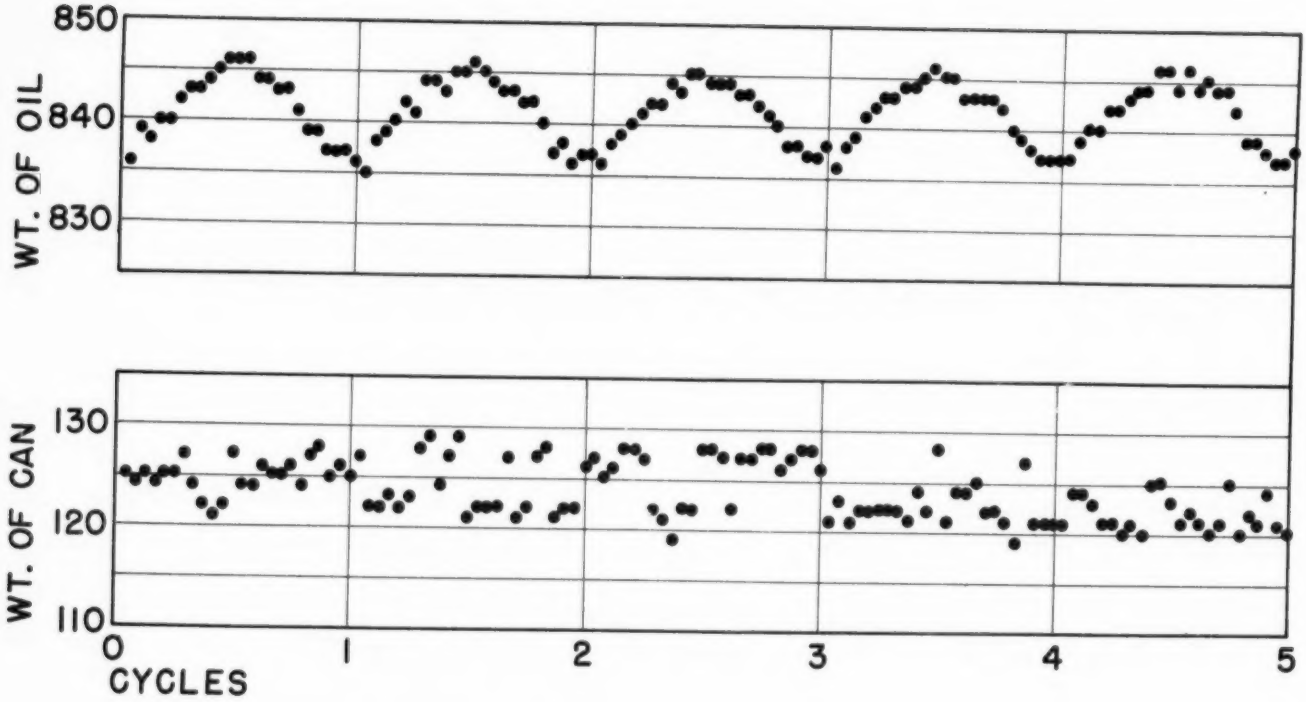


Fig. 3. Original Operation On SAE 30 Oil.

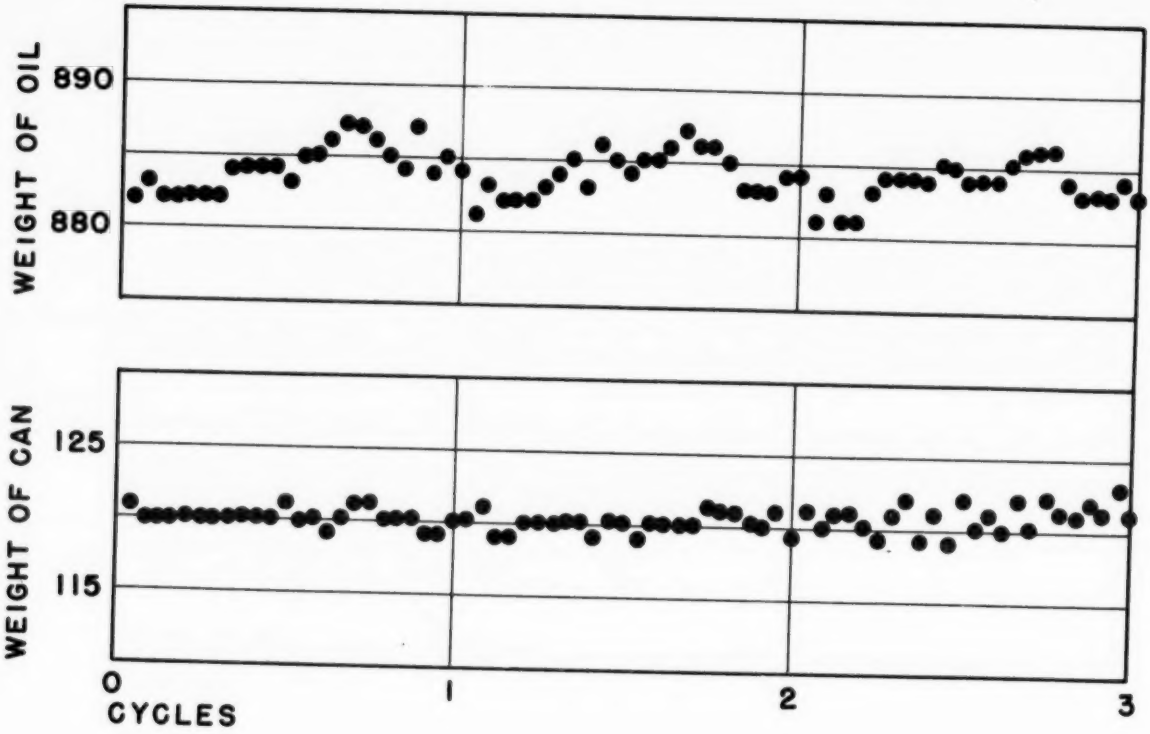


Fig. 4. Operation After Adjustment On SAE 30 Oil.

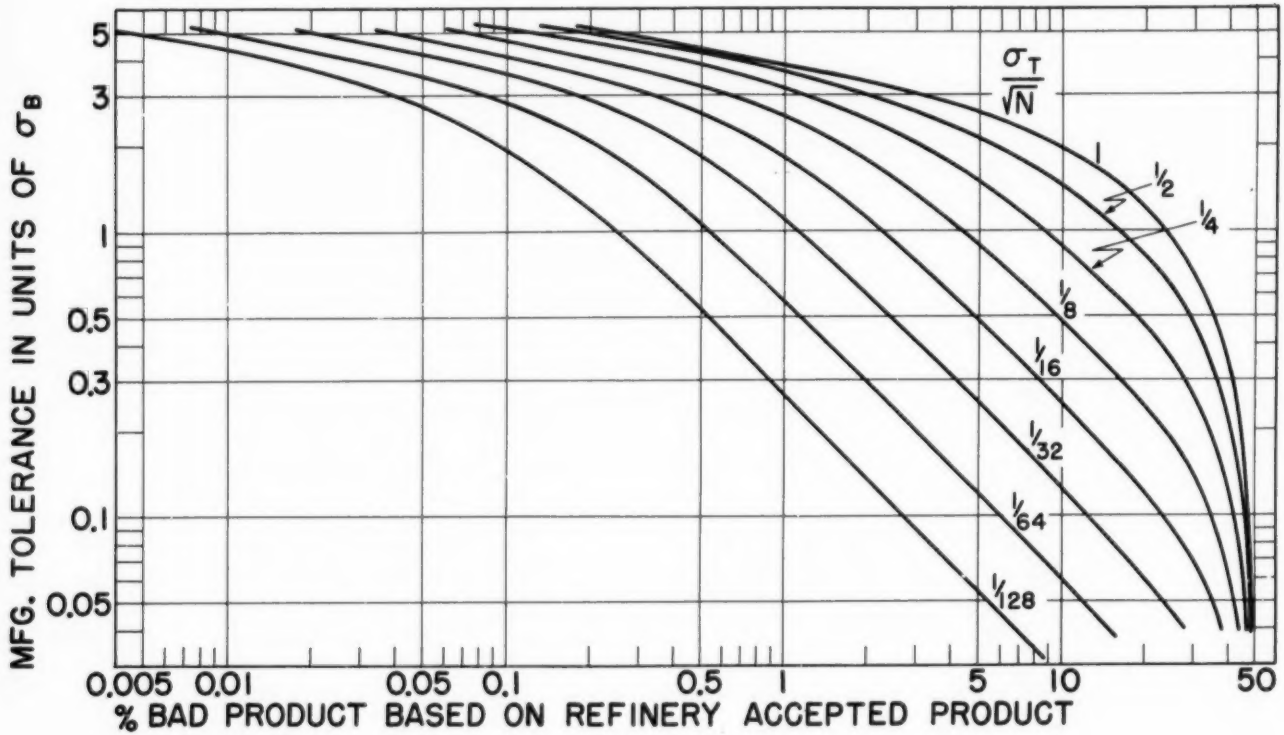


Fig. 5. Effect Of Variables On Bad Product.

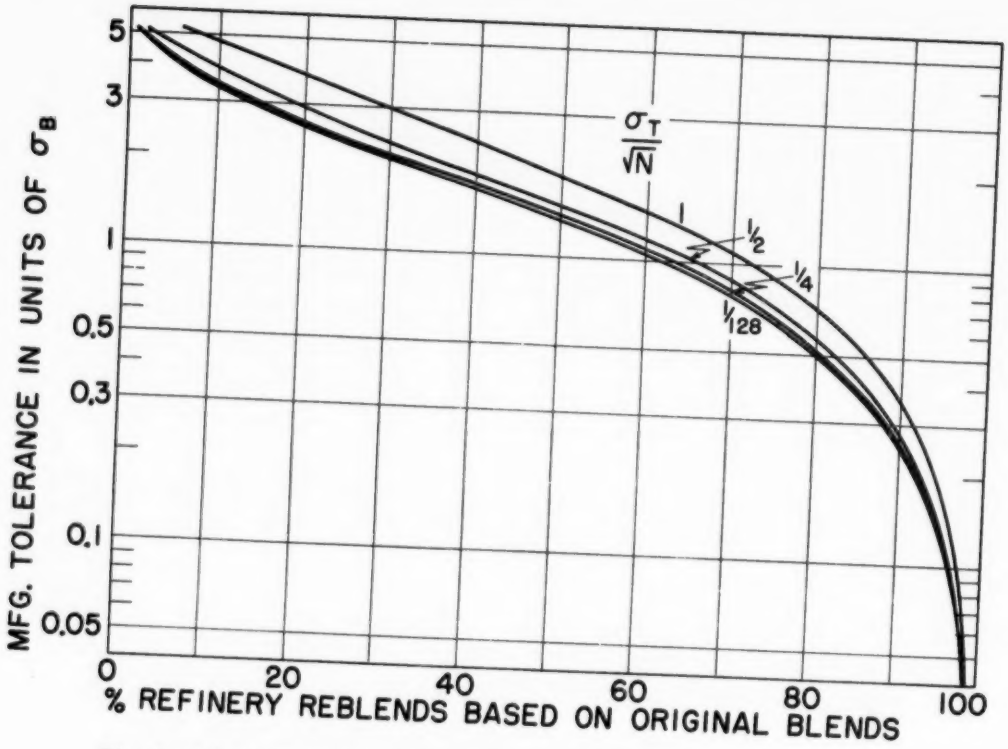


Fig. 6. Effect Of Variables On Refinery Reblends.

WORK SAMPLING (RATIO DELAY STUDY) TECHNIQUE

C. L. Brisley
Wolverine Tube Division

You honor me very greatly by asking me to present the Work Sampling or Ratio Delay Study technique before your Seventh Annual Convention of the American Society for Quality Control. The subject of sampling is one with which you are all very well acquainted. Work Sampling, or the Ratio Delay Study, is a random sampling method of getting facts about machines and human activities.

Harry Lee Waddell, Editor of FACTORY MANAGEMENT & MAINTENANCE, introduced the term Work Sampling for this new concept. This term, from a psychological point of view, places the emphasis on the word "work" rather than on the word "delay" as in the case of Ratio Delay.

We at the Wolverine Tube Division, therefore, prefer the term Work Sampling, so I'll be using this term in place of the Ratio Delay term which is better known in the field of Industrial Engineering.

First, I shall give you a little background on the Ratio Delay Study technique. It was first introduced in this country by Professor Robert Lee Morrow, who was then with New York University. He presented a paper on the subject before the 1940 Annual Meeting of the American Society of Mechanical Engineers. In the April, 1941, issue of MECHANICAL ENGINEERING, the following excerpt is taken with respect to this paper:

> "In searching for a better way to determine delay and variation allowances, a method used by L.H.C. Tippett, in the English Textile Industry, described in the JOURNAL OF THE TEXTILE INSTITUTE Transactions, February 1935, was put to practical tests in three separate industries, by graduate time-study students, under the direction of the Department of Administrative Engineering at New York University.
>
> "These studies, the results of which were reported in Mr. Morrow's paper, show the new method to be a statistical technique, highly practical, and productive of more accurate results at less cost than methods previously employed. As every operation, in every industry on which cost or wage-payment standards are set, must include the allowances mentioned, the importance of this new method is evident. The new method is called 'ratio delay study'."

As a result of Professor Morrow's work and the experience of L.H.C. Tippett in the English Textile Industry, a number of companies have made use of this approach. Perhaps your company is one.

Many who have used the Ratio Delay technique have envisioned many more uses for this concept. The random observation method applies to any activity on which all the facts are needed to understand what is happening and to find out where the opportunities for cost reduction lie. Ratio Delay actually describes only a part of the broad use of the tool. We feel that Work Sampling is a broader term for studies beyond that of analyzing delays.

WHAT IS THE TECHNIQUE?

The Work Sampling technique is based on the laws of probability. With this Society, it is not necessary to explain these laws by resorting to gambling games such as shooting of "craps", the tossing of a coin, or poker. You are all familiar with the normal distribution curve and the laws of probability.

Nevertheless, the same general laws of variation that govern statistical quality control may also be discovered in the MAKE READY, DO, and PUT-AWAY parts of any operation.

The work sampling study is a sampling method of evaluating the possibilities that an operation will be performed in a prescribed manner. Just as in the case of throwing dice, one score may not necessarily occur as frequently as another because of the law of chance. In the case of the dice, we can very easily pre-determine what the laws of chance are. In the case of production operations, we have to throw the dice a good many times, or take a sample to find out what the possibilities are. We must have some method of obtaining a sample from which to determine what a group of people or machines may be doing.

A work sample comprises a selection of a number of elements of work from the total number of elements. If these samples are chosen in such a manner that one element of work has as good a chance of being selected as another, we have what we call a random sample.

A simple illustration will present the theory of the work sampling study. Supposing we break a job down into four parts: (SHOW SLIDE) Set-Up, Operating Cycle, Maintenance and Personal time. In order to illustrate, let us assume that we select just one machine and we make ten observations at random intervals during the day. At the end of the day our tally will be as follows:

Element	Tally	Ratio	Percent
Operation	𝍸 /	6/10	60
Set-Up	//	2/10	20
Maintenance	/	1/10	10
Personal	/	1/10	10

This gives us a ratio as compared to the total of 6 to 10; 2 to 10; 1 to 10; and 1 to 10. Or, converting these ratios to percentages, we have 60%; 20%; 10%; and 10%. This is an example of studying a machine operation.

Let us analyze an engineering job. (SHOW SLIDE) Here again, we break a job down into four elements: Receiving Instructions, Drafting, Filling Out Engineering Orders, and Personal, and here we have tallies at the end of the day as we did in the first case.

Element	Tally	Ratio	Percent
Receiving Instructions	//	2/10	20
Drafting	𝍸 /	6/10	60
Filling Out Eng'r Orders	/	1/10	10
Personal	/	1/10	10

The ratios run 2/10, 6/10, 1/10, and 1/10. And the percentages run 20%, 60%, 10% and 10%.

The theory is that if we multiply 20% times 480 minutes that these people may work during the day, we would arrive at 96 minutes for Receiving Instructions in the case of the engineering study or 96 minutes for Set-Up in the case of the machine study.

Obviously, we have made only 10 observations and the percentage of error would be quite great. But, if we increase these 10 observations to 100, 1,000, 5,000, or even 10,000, you can see that the percentage of error would decrease. We are going to prove that by another demonstration in which you are all going to participate.

(PASS OUT WORK SAMPLING EXPLANATION, "A HOSPITAL LAUNDRY", "WORK SAMPLING HOSPITAL LAUNDRY OBSERVATIONS SHEET", WORK SAMPLING GRAPH TO DETERMINE LENGTH OF STUDY. ALLOW FORTY MINUTES BROKEN DOWN AS FOLLOWS: EXPLANATION - 10; FILM - 20; CALCULATIONS - 10)

You are going to make a Work Sampling study in this conference, of a hospital laundry operation of pressing uniforms. In the Conference Room you are going to simulate a work sampling study just as if you were in the laundry passing this operation at random intervals.

RANDOM NUMBERS:

Figure 1 on the next page is a work sampling form on which is listed fifty random numbers placed in sequence. As the film is projected a wire recorded count will be played simultaneously, giving you ten counts per minute.

You should make a tally in the tally column for the numbers which you have on your Random Number Table for the element of work which occurs the instant your number comes up (as shown on Figure 1).

As an example: Suppose your first five numbers are:

3, 11, 12, 16, and 18

No. 3 might be observed at the instant the operator is reaching for the operating button of the No. 1 Mushroom Press. You would make your tally after Element A.

Nos. 11 and 12 may come while she is adjusting uniforms on the Body Press. If so, two tallies (//) should be placed after Element C. No. 16 may come while she is spraying on the Mushroom Press. If so, make a mark after Element D. Spraying on any press should be recorded under Element D, should it occur the instant your number is called.

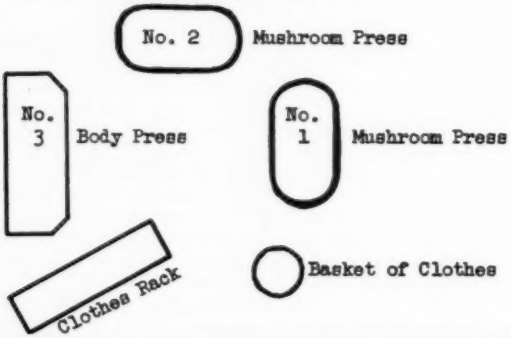
No. 18 may come as she is picking up another garment for the No. 3 Body Press. This garment has been partially pressed on the No. 1 or No. 2 Mushroom Press, and placed on the table beside the press for final pressing on the No. 3 Body Press.

FIGURE I

WORK SAMPLING

A HOSPITAL LAUNDRY

WORK PLACE LAYOUT



RANDOM NUMBERS		ELEMENT OF WORK	ACTUAL %	TALLY	TOTAL	%
7	104	A. Adjust Uniform on #1				
14	105	Mushroom Press	22.9	𝍸 𝍸	12	24
16	106			//		
19	109					
20	114	B. Adjust Uniform on #2				
26	127	Mushroom Press	21.7	𝍸 𝍸	//	22
27	129			/		
30	132					
32	135	C. Adjust Uniform on #3				
39	137	Body Press	36.2	𝍸 𝍸	17	34
43	138			𝍸 //		
50	141					
54	142	D. Spray Uniform	6.4	///	3	6
55	144					
56	147	E. Dispose of Completed Uniform				
64	149	from #3 Body Press to Rack				
68	151	and Get Another From Mush-				
69	152	room Press	6.7	///	3	6
77	155					
80	158	F. Dispose of Uniform beside				
82	159	#1 or #2 Mushroom Press and				
89	162	get another uniform from				
90	173	Basket	4.5	//	2	4
95	174					
97	187	G. Get bundle of uniforms and				
		place in basket	1.6	//	2	4
		TOTAL	100		50	100

INTERPRETATION OF NUMBERS:

For this demonstration the interpretation of numbers like 93 or 152 is to record the element of work on the last digit as the narrator calls ninety-three or one hundred fifty-two.

(SHOW FILM - 20 MINUTES)

ACCURACY OF WORK SAMPLING STUDY:

You can determine the accuracy of your tally with the actual percentage which has been determined very accurately by micro-motion study.

HOW TO DETERMINE PERCENTAGE OF TIME:

Assuming that you have a total of fifty observations and element tally marks (~~𝍸~~ ~~𝍸~~ //) for element of work "A"; thus

$$12 \div 50 \times 100\% = 24\%.$$

Compute all of your percentages in this manner.

HOW DO WE KNOW THAT THIS PERCENTAGE IS ACCURATE?

Since this operation has been filmed, we can compare our percentage with the micro-motion study frame count. The film passed through the motor driven camera at a constant rate of 1000 frames (pictures) per minute. A micro-motion study involved counting these pictures for each element. For example, for Element "A" there were

4534 frames ÷ the total of 19,744 frames x 100% = 22.9%

Since the film was taken at 1000 frames per minute, the picture is exactly 19.744 minutes long. Therefore,

Element "A" = 22.9% x 19.74 min. = 4.52 min.

SUPPOSE YOU DON'T KNOW THE CORRECT PERCENTAGE ON A STUDY YOU MAY MAKE?

You can determine the percentage of error by a Statistical Standard Error Formula, $\sigma = \sqrt{\frac{pq}{n}}$. In this meeting we will not discuss this equation.

HOW TO DETERMINE THE LENGTH OF STUDY:

A simple and empirical method, however, may be demonstrated by a graph to determine the length of study. Let us consider Element A and let us assume that each person's 50 observations represents a day. We shall plot a graph to determine the length of study.

(PICK OUT TEN PEOPLE AT RANDOM AND ASK EACH PERSON HOW MANY TALLIES HE HAS ON ELEMENT "A", ASSUMING THAT THEY MAY BE AS IN FIGURES 2 AND 3.)

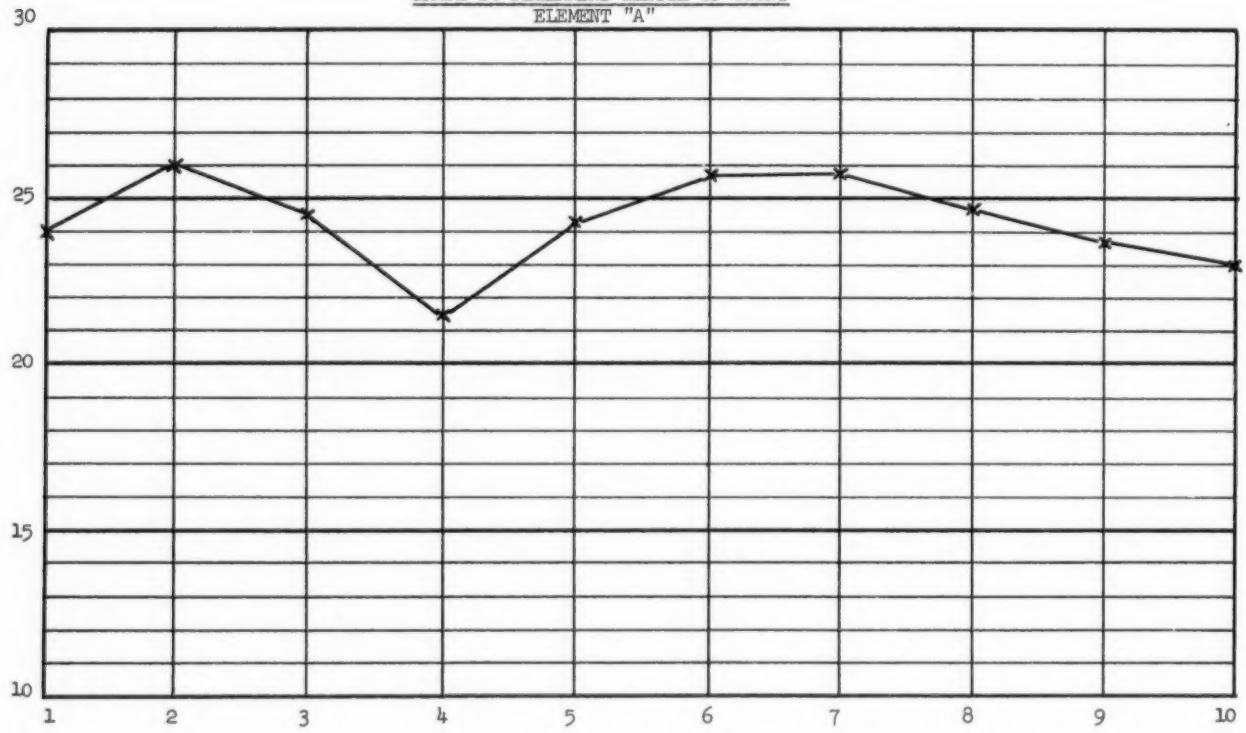
We shall plot a graph to determine the length of study. This

FIGURE 2

ACCUMULATIVE PERCENTAGE BASED ON ACCUMULATIVE NUMBER OF OBSERVATIONS COMPARED TO ACCUMULATIVE TOTAL FOR ELEMENT "A"

	Days	1	2	3	4	5	6	7	8	9	10
	Tally Element "A"	12	15	10	6	18	17	12	9	7	9
	Total Daily Observations	50	50	50	50	50	50	50	50	50	50
(1)	Accumulative Total Observations	50	100	150	200	250	300	350	400	450	500
(2)	Accumulative Tally Element "A"	12	27	37	43	61	78	90	99	106	115
(3)	Ratio: Accumulative Tally Element "A" to Accumulative Total	$\frac{12}{50}$	$\frac{27}{100}$	$\frac{37}{150}$	$\frac{43}{200}$	$\frac{61}{250}$	$\frac{78}{300}$	$\frac{90}{350}$	$\frac{99}{400}$	$\frac{106}{450}$	$\frac{115}{500}$
(4)	Accumulative Percentage	24.0	27.0	24.6	21.5	24.3	25.8	25.7	24.7	23.6	23.0

FIGURE 3
WORK SAMPLING
GRAPH TO DETERMINE LENGTH OF STUDY
ELEMENT "A"



graph comprises plotting the accumulative percentage of each element on a chart of percentage vs. the number of observations. Let us use as an example, Element A, "Adjust Uniform on No. 1 Mushroom Press". The number of observations for each of ten days is shown in Figure 2, "Accumulative Percentage Based on Accumulative Number of Observations Compared to Accumulative Total for Element "A".

The computations are as follows:

A. (2) "Accumulative Tally of Element"A" divided by (1) "Accumulative Total Observations" equals (3) "Ratio Accumulative Tally Element A to Accumulative Total"

B. (3) "Ratio Accumulative Tally Element "A" to Accumulative Total" times 100% equals (4) "The Accumulative Percentage".

From the (4) "Accumulative Percentage" a graph is made similar to that shown in Figure 3, "Work Sampling Graph to Determine Length of Study". By the picture presented by this graph (Figure 3) it is very easy to see that the percentage is now running, after ten days' study, between 21.5% and 27%. During the last six days, the range has been 23% to 25.8%.

We observe that the accumulative percentage curve of Element "A" is approaching a straight line. With more observations, our accumulative percentage approaches the true average percentage of time expended on this element. That is, our percentage of accuracy increases with more observations. Because the percentages are close together during the last six days in Element "A", it is reasonable to assume that the percentage of accuracy is sufficient for most purposes.

Using the percentage arrived at on the tenth day for Element "A", and since the operator works an 8-hour day or 480 minutes per day, the time expended on Element A, "Adjust Uniform on No. 1 Mushroom Press", is 23% x 480 minutes = 110 minutes per day, or multiplying 23% x 60 minutes, the operator is averaging 13.8 minutes per hour on this element of work.

In order to simplify the example, we've only illustrated observations on one element. Let us assume that a laundry has ten similar operations such as this. While the observer is making his fifty trips he can just as well observe ten operations as he can one. Obviously, it would take a little longer. If he did observe ten similar operations, he would have 500 observations each day. In ten days, he would make 5,000 observations and the percentage of accuracy would be much greater.

APPLICATIONS OF WORK SAMPLING:

Now the question is, "Where can we use this Approach?" At the Wolverine Tube Division, we have used this technique to determine time study allowances. If you use time standards for any purpose, you generally add allowances to the basic work time for "miscellaneous work" and "necessary delays". The general practice is to make continuous time studies over a period of time long enough to get a typical

cross section of the process conditions. Continuous time study takes a lot of specially trained manpower, but you must have factual data to establish reasonable allowances. Now, we get the same data by work sampling and it requires only a fraction of the man-hours needed to make continuous time studies. Observers require much less training than a qualified time study man. Experience has taught all of us that the data are sufficiently accurate.

Decatur, Alabama Plant:

When our new Decatur, Alabama, Plant began operating in 1948, it was equipped with the most modern machines. We had many production problems in the beginning. Inexperience with the new equipment caused some "bugs" and needed improvements in the equipment caused others. In fact, we had so many problems, we hardly knew where to start. We needed something to tell us quickly of the location and the relative magnitude and causes of our more serious problems. Experience with finding delays in our Detroit Plant by work sampling suggested using the technique on our new plant problems.

Random observations instead of continuous observations permitted us to get the facts about many operations at once. The facts told us where and why we weren't using our potential production capacity. We concentrated all our problem solving attention on the worst cases first and the others later. Through our Work Simplification Follow-Up Program we were able to improve methods and we made a great many work sampling studies at intervals of approximately every two or three months to check up on our progress and as a result of continually working on those items which proved to be the greatest problem, we were able to increase our productivity and decrease our delays.

Harper Hospital:

This technique has been applied at Harper Hospital in Detroit. About two years ago, the management of Harper Hospital came to Wayne University and to the Detroit Chapter of the Society for the Advancement of Management for help with respect to their problems. As a result, a joint committee was set up comprising members of the S.A.M. and Professors of Wayne University to help the hospital in the solution of these problems. It was apparent that the higher echelons of the nursing personnel were doing some jobs that lower skilled personnel could be doing. The administration of the hospital were anxious to get at the bottom of this problem because of the shortage of nurses. We suggested that they make use of the Work Sampling technique. As a consequence, I spent some time with three hospital administrative understudies and showed them how to make a work sampling study. The first study showed that Head Nurses were spending about 20% of their time charting, a job for non-supervisors. The study also showed that registered nurses spent much time running errands for patients and performing menial tasks. Reassignment of duties gave the professional work to the nurses and other work to nurse's aides, orderlies and messengers. The result has been that the same force of nurses now care for more than twice as many patients in a test unit now set up at the hospital. This program has had far reaching influence.

Shortage of Engineers Problem:

The problem with respect to the shortage of engineers is quite parallel to this shortage of nurses problem. From April 17th to May 2nd, 1952 at our Decatur, Alabama, Plant, the work sampling technique was used in analyzing engineering work. The Plant Engineer is very alert to efficient methods and the study reflected his effectiveness. The approach was as follows:

We first consulted with the Plant Manager who gave us permission to take up the proposition of a Work Sampling study with the Factory Manager. When the Factory Manager was approached, he called in the Plant Engineer. In this meeting, the Supervisor of Industrial Engineering at the Decatur Plant and I presented the general approach and reminded the Factory Manager and the Plant Engineer of the success we had had with the work sampling technique on production. The Harper Hospital approach was also emphasized. We pointed out that it was our opinion that the shortage of engineers and the shortage of nurses problems were parallel. We, therefore, suggested that a work sampling study be made in the Engineering Section. The Plant Engineer readily agreed and set a time for us to meet with him and his engineers. At the meeting the engineers were fully cooperative. It was pointed out to them that the whole program was designed to enhance their jobs and to take away from them clerical work and menial tasks if possible. It was also pointed out that they would participate in the decisions that were made.

After the 14-day period, 252 trips had been made. The average trip was approximately 25 minutes. Each draftsman spent 26 minutes per day walking. As a result of this study, the following decisions were made for the purpose of utilizing the engineers' time more effectively:

1. Consolidate trips to the plant.
2. Make use of bicycles to reduce walking time.
3. Stores Section should train someone to check obsolete items.
4. Telephone could be used more.
5. Use dictating machines instead of taking notes.

Executive Study:

As far as I know, no such study as we have made on our executives, has ever been made before. We were in somewhat the same position as the little boy of a friend of ours who was drawing a picture on his blackboard and his mother said, "Who are you drawing?"

The little boy said, "I'm drawing a picture of God."

His mother said, "That's silly. No one know what God looks like."

The little boy said, "Well, they will know when I get through."

In the Detroit Plant, beginning July 21, 1952, we started making a study of our executives. Because of the confidential nature of many of the executives' work, I made this study myself. I made 10 random trips daily and observed what each executive was doing as I passed his

office. The purpose of this study was to give our executives factual information whereby they could improve in the utilization of their time.

In conclusion, it is my opinion, as a result of the work that we have done at the Wolverine Tube Division, that the work sampling technique can be a very valuable tool for any business no matter what that business may be. We are all interested in the utilization of time.

I remember about six years ago, when I first came to Detroit, I visited Greenfield Village and upon visiting the office of Thomas A. Edison in His Menlo Park Laboratory, I was very much impressed by the fact that Henry Ford had removed the face of the clock to commemorate the philosophy of Thomas Edison, which is:

> "Time is not measured by minutes and seconds, but by accomplishments."

I hope that you will use the work sampling technique to accomplish more in your organization.

QUALITY CONTROL'S IMPACT ON THE USE OF THE STANDARD INCH

John A. Harrington
The DoALL Company

The use of quality control methods of inspection has made necessary better calibration of primary measuring instruments and has compelled inspection departments to standardize the inch which they are using. Let me call your attention to the fact that there are still, however, several thousand different inches being used by industry in this country and needless to say this causes considerable confusion and expense, particularly when components from various plants have to be assembled. Even within some plants we will find a variation in the inch used in the different departments. This lack of use of the standard inch contributes considerably to the expense of production, for when different plants or departments are using their own inch and these components eventually reach the assembly line, needless expense is incurred and re-measuring or re-machining of these parts is necessary, providing they can be reused. This expense is of considerable significance for if we take into consideration the average rejects that are produced in this country today, due to lack of proper quality control, we will find that it approximates $500,000,000 per year in material and the labor of 300,000 mechanics. This is well worth saving and it is for this purpose that we would like to discuss the standard inch as it pertains to quality control and a few of the methods that will help accomplish this. We would like also to take this opportunity of impressing upon the audiance the necessity of recognizing and using the standard inch if this saving is to be accomplished.

We have in this country a standard inch which has been legalized by an act of congress and which is the inch we are supposed to use, but unfortunately many of us fail to understand what an inch really is and generally use our own inch. I doubt if there are many of you present that can define a standard inch. This is rather startling because most of us earn our bread and butter interpreting this inch into articles of production.

About 1793, just after the French Revolution, the French Government put into effect a system of dimensions, volumes and weights known as the metric system and this system was based upon the length of a metal rod called a meter upon which there was scribed two lines, one at each end. The distance between these two lines was supposed to have been one ten-millionth of the quadrant of the earth on a line stretching from the North Pole through Paris to the equator. This standard served very well for a number of years but eventually it was found that replicas of it were almost impossible to reproduce with the necessary accuracy and when taken to other parts of the world the measurements failed to agree. This lack of accuracy was not due entirely to temperature changes for they could be compensated for, but there were other factors entering into the picture, such as the pull of gravity and metallurgical changes in the metal itself which presented many difficulties in maintaining the exact dimension. The necessity for being able to duplicate these dimensions became increasingly important for upon the accuracy of the dimension rested the accuracy of volume and weight. As you undoubtedly recall, a unit of volume is a cube of a specified dimension and the unit of weight is also a cube of distilled water at a standard temperature, the cube being of a specified dimension. So you see that without an accurate dimension, standardization of weights, measures and lengths in various parts of the

world would be impossible.

Prior to 1890 Albert Michelson, who was then a physicist at the University of Chicago, had established the fact that the wave length of light, as emitted from the various elements, was always of a uniform length for a particular color. In 1892 he was called over to Paris to determine the length of the then existing meter at the International Bureau of Weights and Measures. Together with physicists from other countries he worked for more than two years and determined that the length of the meter was 1,553,164.13 wave lengths of cadium red. This number of light waves was checked and rechecked a number of times by the various physicists throughout the world and became the standard length of all existing dimensional units used by the various countries.

In this country the legal inch is 25.4 millimeters, the length of which is determined by 39,450.3689 wave lengths of cadium red. The accuracy of this unit of measurement can be better appreciated when we realize that light of certain color traveling a distance of 92 million miles from the sun and whose wave length is approximately 12 millionths of an inch is so accurate that it never becomes out of phase so as to cause interference, even though there are this vast number of duplications of the wave form; so you can see what an accurate natural standard we have access to. By the use of an interferometer we can set up the interference bands from our light source at any time in any place and calibrate in terms of the standard inch.

In the calibration of our commonly used measuring instruments we are generally concerned with very small differences, seldom more than a thousandth or so of an inch, and most of the time with measurements considerably smaller. It is not necessary for us to have expensive types of interferometers to do this checking. Industry generally uses a simplified form of the interferometer commonly called a monolight and we use helium as our light source. This, together with a few optical flats made of fused quartz, gives us the same accuracy at considerably less expense in equipment. This equipment does not measure as large differences as the laboratory type of interferometer, which in some cases can measure direct as much as 8" or more.

The monolight just referred to is a very essential piece of equipment in an inspection department for without it, it becomes extremely difficult to keep our measuring equipment calibrated to the tolerances required. This primary calibration instrument is not expensive. Very satisfactory monolights and flats for our work can be purchased for two or three hundred dollars as compared to ten or fifteen thousand dollars for the more expensive interferometers. Without the monolight and optical flats, it becomes necessary for most inspection divisions to frequently send their various measuring instruments, such as micrometers, height gages, plug gages, thread gages, gage blocks, etc. to some qualified and capable calibrating service for proper calibration. This is time consuming and often requires unnecessary duplication of the various measuring instruments, while if the inspection department is equipped with its own monolights, optical flats, and calibrated gage blocks, it can do most of the calibrating required for quality control.

Now let us see what sometimes happens when we fail to keep our measuring instruments standardized. I would like to cite a recent experience as an example of what sometimes happens. A very large plant was setting up pilot operations for the production of jet fuel pumps, and after con-

siderable research, investigation and collaboration of the Air Force, it was determined that the initial objective should be a maximum of 4% defectives and the method to be used in production was worked out on this basis. After several runs of considerable size had been made, agreement was reached with this percentage when using the same measuring instruments which had been used in the initial sampling and it was decided to proceed on this basis, since the possibilities of considerable improvement existed, and it was believed that the defective percentage could be reduced to about 2% as soon as the personnel had been more thoroughly trained and became accustomed to the operation. But after the first run had been made and accepted by the quality control department of the contractor, the air ordnance inspectors stepped in and found that the rejection percentage was closer to 7% and therefore unacceptable. We were called in about this time. We found that the original standard diviation or sigma, when using their measuring equipment, was 1.73, but taking a lead from the ordnance inspectors we then proceeded to calibrate the measuring equipment which they had used and found that there was considerable variation in the various instruments from the standard inch and when the sigma was recalculated, based upon the standard inch, the rejection percentage calculated 6.92%. The lack of appreciation of the standard inch in this case had caused an error of considerable magnitude, approximately 75%, which I think you will agree is too large an error to be tolerated when using statistical methods.

It would not do much good to call attention to a condition such as this without supplying the information necessary to correct it and to show some of the reasons why these errors occur. Now, since our time is rather limited, I will only mention a few of the more common mistakes which cause this diversity of inches.

First let me refresh your memory concerning the tolerances and clearances for our standard of the measuring instruments we commonly use. Let us take as our example a plug gage. Now in plug gages the tolerance permitted on the gage is about 10% of the tolerance of the part to be measured and let us say that we are using a half inch plug gage of either the solid or air type to measure the bore of a fuel injection pump which has a unilateral tolerance of one ten-thousandth of an inch. This is a rather large tolerance as most of the bores of this type of pump have a tolerance of about 20 millionths of an inch, but we can still use as our example a tolerance for the component of one ten-thousandth of an inch. Now following the rule of the gage makers tolerance we find that the tolerance on the plug gage will be ten-millionths of an inch. This is the gage makers tolerance and we are not going into such things as wear allowances at this time, but only concern ourselves for the moment with the gage makers tolerance which is ten-millionths of an inch. Knowing that many things can happen to steel gages such as warpage, lack of stability, wear, etc., we generally calibrate these gages prior to their initial use and generally use the following procedure. We set up a comparator using a flat serated anvil and place on the anvil gage blocks of the proper dimension and then set the indicator on the comparator at zero. We then roll the plug gage on the anvil under the spindle noting the high reading, and from this high reading determine whether the plug gage is within tolerances or not. I think you will agree with me that this is about the most commonly used method of calibrating plug gages. What I wish to call to your attention is this: we have neglected to take into consideration the toughness of the air film which is trapped between the anvil and the gage block surfaces with a result that we never know whether we are in intimate contact with

these surfaces or not, but we do know that chances are all against us having a correct measurement. Now this air film can cause a considerable error for air, when it is trapped between two flat surfaces, is rather difficult to eliminate. The amount of this error we can readily prove. We have here a comparator such as is generally used for this type of measurement together with a serated anvil and a gage block. We will ask one of the gentlemen here to place this gage block on the anvil in the same manner he generally does in setting up a comparator. Now we will ask him to duplicate this in the same manner, only this time he will put an optical flat on top of the comparator anvil and he will pretend, for the moment, that the surface of this flat is the surface of the anvil on the comparator. I will now pick up the gage block and the optical flat and after reversing it, place it under the monolight so that we can see the number of fringe lines or in other words, the thickness of the air film trapped between these two surfaces. We will do this several times, each time recording the number of interference bands and since we know that the interference band of helium is 11.6 of an inch we are now able to accurately determine the thickness and variations of this thickness which we have used in setting up our comparator, thereby making our dimension larger to this extent than the standard inch. You will note that if we multiply out these fringes by 11.6 millionths that several of them amount to considerably more than the tolerance permitted--in one case--close to two ten-thousandths of an inch, so that instead of having a tolerance on our plug gage of ten-millionths of an inch we have two hundred millionths of an inch, or our tolerance measurement is 20 times more than it should be. This, gentlemen, is one of the frequent causes of erroneous measurements and contributes to the diversity of inches which industry is now using. It is one of the most common causes of a lack of agreement between the contracting inspectors and ordnance inspectors for each of them often make this same mistake and as you have seen, it would be impossible for them to agree on their measurement. A simple way in which this cause of error can be eliminated and our measurements be made accurate is by using a 3 point anvil such as is illustrated here:

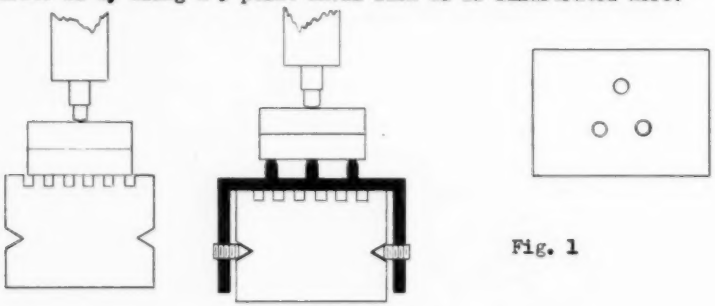


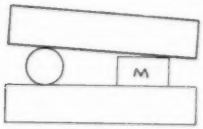
Fig. 1

Now since the contact with the flat surface is only 3 points the unit pressure is sufficiently high to penetrate the air film that we have just demonstrated and we know that we have contact between the two surfaces. This can be readily proven by removing the block and repeating our measurements, for so long as we are careful not to introduce any heat, you will notice that you are able to repeat your measurements. We were unable to do this on the flat anvils.

This does not, however, answer our problem of calibrating the plug gage. In order to calibrate this gage properly, place the plug gage on an

optical flat alongside of a gage block of the nominal dimension of the plug gage as illustrated in Fig. 2.

As we demonstrated before, we are now in a position, by reversing the flat and looking through the bottom, to see if we have intimate contact and if there is any air retained between the gage block and the optical flat. If any air is present, we can eliminate it by wringing until no bands or color shows and we then know that our dimensions will be correct. Now, by placing another optical flat on top of the gage block and the plug gage, as illustrated—



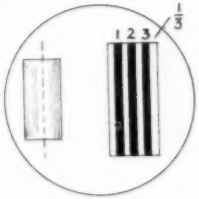


Fig. 2

we are able to read in millionths of an inch the exact diameter of the gage, and by rotating it we can determine its roundness. Now since we know the dimension of this plug gage it is in a sense a master, since any gage that we know the exact dimension of can be used for setting up comparators or other instruments, and allowances can be made for any variations from the exact size.

We can now use this plug gage for setting our comparator, for by rolling the gage under the spindle we only have a line contact with the flat surface, which readily displaces the air film and we are assured of the correct setting. However, if we have a number of round parts to inspect, it is then better to use a 4 point "V" block, such as is illustrated here,

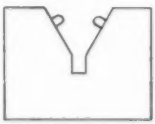
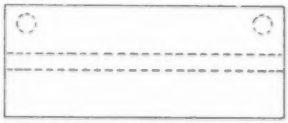


Fig. 3



for it is a more rapid method and when properly used, insures greater accuracy.

In checking round parts such as plug gages or high speed shafts, which have close tolerances, there is one mistake that is often made which at times causes considerable confusion. It is the use of the more common type of "V" block placed under a comparator. We rotate the part and record the variation in the dimensions as the amount by which the part is out of round. Now let us see actually what happens.

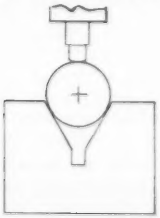
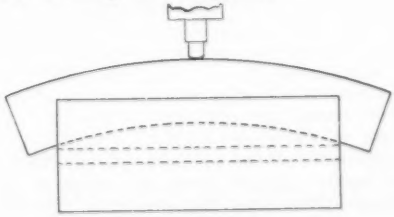


Fig. 4



You will note that as we rotate this part under the comparator and in the "V" block, that if there should be any warpage whatever, I am not recording the out of round dimensions necessarily. I may also be introducing the error of warpage so that my dimension may not be correct and as a practical matter, very often isn't. This source of error can readily be eliminated by the following procedure.

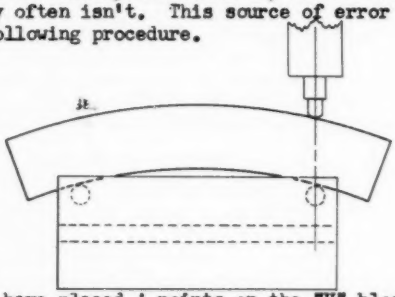


Fig. 5

You will note here that we have placed 4 points on the "V" block and that we have swung the comparator head so that it is in alignment with one set of points. Now if I rotate the part, I am reading its roundness, and even though there is a warpage I do not record it. In the next illustration—

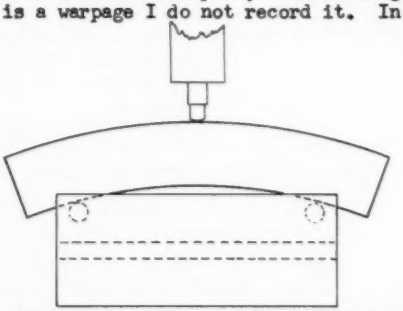


Fig. 6

—you will note that we have swung the comparator head over to a point midway between the supporting point and now if I rotate the part I will record the warpage. With this information I am in a position then to state exactly what the dimensional conditions are for this particular part and the information that I pass on as an inspector will be correct.

Undoubtedly the question has come to your mind as to the air film which may be present when using a stack of gage blocks and if you have reached the conclusion that there is likely to be an air film between these gage blocks, you are correct, so the question remains as to how we can insure accurate dimensions when using a stack of gage blocks.

It has been found, and you can readily prove it yourself, that it takes about 40 lbs. per square inch of pressure to be sure that you have displaced the air film between two gaging surfaces, such as the U.S. Standard gage block which is rectangular and has about 1/2 square inch of contact area. This can be proven with the use of an optical flat by placing the gage block on the flat and keep adding pressure until all fringe lines and color have disappeared. This is generally done by placing a mirror under the optical flat so that you get the reflection from the bottom of the gage block. To insure the correct amount of pressure, this arrangement, which is part of the Micro-Step System, has been so designed that normal finger pressure on these clamping knobs exerts about 50 lbs. maximum, depending upon the strength of the operator. It will vary for

different people from 40 to 50 lbs. You see that by the use of this system that if I clean my gage blocks and then place them in the holder, I do not have to wring the blocks together and as you will find after I have clamped them, I can unclamp the holder and find that my gage blocks are wrung together and I also am assured of intimate contact of the surfaces and therefore of correct dimension. This, of course, is based upon the assumption that the gage blocks have been kept calibrated, therefore are flat and of a known dimension. As you can readily see, this method of using gage blocks will greatly prolong the useful life of the gages as there is a minimum amount of wear, since there is no necessity of wringing these blocks together and I will now also be able, with assurance, to calibrate any of my measuring instruments that are used for inspection and sampling and I will not be transmitting inaccuracies to other measuring devices. Therefore, when I compute the standard deviations of parts as they come off the production line I will know that I am conforming to the United States Standard Inch and will not be fostering on the public an inch of my own which bears little relationship to the standard inch I am supposed to be using.

There are other devices on the market, such as clamps and screws for holding together a stack of blocks, but unfortunately none of these, with the exception of the Micro-Step System, are so designed that they will insure accuracy. As an illustration I might cite one of the commonly used square types of blocks which are clamped together by a rod through a center hole. If you will accummulate a stack of these blocks approximately 10" and have them clamped by several people in your department, you will note that the pressure which they use will shorten the stack various amounts up to and beyond 12 millionths of an inch per inch. From this we conclude it is only through properly designed holding fixtures that you have the assurance of correct dimensions.

In summing up the impact of statistical control on inspection in production we find the following: first, statistical methods are gradually forcing upon industry the necessity of standardizing their inch. Second, by standardizing the inch we find that better calibration of our measuring instruments is necessary. Better calibration is leading us to a better understanding of our measuring instruments which, in turn, brings us back to better dimensional control through statistics and more economical and efficient production.

STATISTICAL CONTROL IN CONTINUOUS CHEMICAL PROCESSING

Henry W. Isleib
and
Robert M. Hofstead
Bristol Laboratories Inc.

"Continuous Process" is a term commonly used in industry. By its very nature it is employed most appropriately in modern progressive chemical industries. Since the term continuous process will be used frequently in the following discussion, a definition is presented with the hope of obtaining a basis for mutual understanding.

Defined by Kirkbride (1): "A system consists of the mass and energy that are confined by known boundaries. It may be either static (no movement) or dynamic (in motion). A process, on the other hand, is the method by which a change takes place within the system.

There are two fundamental and distinct types of processes, viz., batch and continuous. In the batch process the mass is charged to the system at one time and the process is carried to completion. In the continuous process the mass is charged to the system continuously and the products are removed continuously."

Examples that may be presented of continuous processes found in the chemical industry range from the huge gas diffusion process in operation at the U. S. Government's atomic energy plant, Oak Ridge, Tennessee, or the large oil refinery operations located throughout the world, to something small and simple as the continuous distillation of tap water in the laboratory.

Continuing a step further, the concept of "unit operations" is introduced. These are operations common to different industries. In effect, Brown (2), suggests that: "Any chemical process, on whatever scale conducted, may be resolved into a coordinate series of what may be termed "Unit Operations," as pulverizing, drying, roasting, crystallizing, filtering, distilling, or extracting. The number of these basic unit operations is not large and relatively few of them are involved in any particular process. The complexity of the treatment of these operations is real and results from the variety of conditions (variables) as to temperature, pressure, or pH under which the unit operations must be carried out in different processes, and from the limitations as to materials of construction and design of apparatus imposed by physical and chemical character of the reacting substance."

Marketable products resulting from the processes encountered in the chemical industries are required, with few exceptions, to fall within ranges of established standards. Obviously, in order for the products to meet these standards, the in-process variables must also be controlled within limits. The in-process variable controls differ in type, complexity, and sensitivity, and these controls give varying degrees of success depending on factors too numerous to mention here.

Utilization of modern and improved control techniques has greatly improved in-process control procedures. In recent years remarkable ad-

vances have been made in instrumentation and measuring techniques; in the accuracy and sensitivity of analytical laboratory tests; and in the uses of the tool known as "Statistical Control."

The success of any in-process control plan still depends on the time, thought and effort applied to it. To bear this out, even with the best controls available; a control plan may not be a complete success because:

A. Valuable controls are omitted which escape recognition.
B. Little effort is made to improve the control plan.
C. Process improvements are not uncovered which would have been indicated through a thorough analysis of the control data.

Statistical control is a unique and exceedingly versatile control technique. It can be applied scientifically to repetitive data, and through the principles of probability and logical application, can provide a streamlined method for separating significant information, pointing out lines of action, and predicting the probable outcome (3).

Statistical control, quality control, and statistical process control will be used interchangeably throughout the discussion and will be considered to have the same meaning.

A comprehension of the basic concepts of statistical control is usually necessary before applying it to data. As an adjunct to this understanding the definitions of the two components of the term quality control are presented.

The term control is defined by Shewhart (4), by stating that: "A phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits."

In analizing quality, Shewhart (5), points out two common aspects: "One of these has to do with the consideration of the quality of a thing as an objective reality independent of the existance of man. The other has to do with what we think, feel, or sense as a result of the objective reality. In other words, there is a subjective side of quality. For example, we are dealing with the subjective concept of quality when we attempt to measure goodness of a thing, for it is impossible to think of a thing as having goodness independent of some human want. In fact, this subjective concept of quality is closely tied up with the utility or value of the objective physical properties of the thing itself."

With these aspects of quality in mind, then, the physical characteristics of the product must be defined to satisfy most nearly the human wants. The human element, therefore, must be of prime consideration in the control of quality since it is part of the basic concept of quality.

Humanics cannot be stressed too strongly - not only does it decide the limits to be placed on the physical characteristics of the product, but it is an integral part of statistical control applications. The

statistical analysis of repetitive measurements is of little value without initiative and imagination to recognize significant information, to point out lines of action, and to predict the probable outcome. Of equal importance is the part it plays in formulating and carrying out the plan. Statistical Control is not purely a mechanical operation. It involves people. Human consideration is the key to the successful application of statistical control.

In a chemical plant the processes, or unit operations, are usually supervised by people with technical training. This technical training may or may not be well supported by knowledge of statistical technique. If the supervisor has a background in statistical control he will be able to use this training to good advantage in a practical control plan. If not, he must rely on the statistician, not only to supply the statistical techniques required for the plan, but also to aid in interpreting the results of the statistical analysis derived therefrom. However, the supervisor in this case is not relieved of the responsibility for understanding the philosophy and objectives of the technique and for using his ingenuity and imagination to discover where it may be applied.

The statistician, on the other hand, must be aware of the implications of the process to contribute to its control. He must be familiar with the properties of materials and limitations of equipment, the chemical mechanisms encountered, the variables which affect the process and the human characteristics involved.

The statistician and technically trained supervisor, working together with all concerned, forms a most effective team to pursue the aims and objectives of statistical process control. These aims being greater process efficiency, maximum yield of acceptable product and reduction of processing costs.

As previously stated, a process is the method by which a change takes place in a system, while a unit operation is one of the basic operations found in the system and may be common to many processes. The various conditions required in a unit operation are dictated by the process. Statistical controls should be applied to the relatively important variables in the several unit operations and statistical techniques used to study the operations collectively to obtain knowledge of their inter-relations and establish overall control of the complete process.

In a continuous chemical process of one or more unit operations, the quantity of a stream entering or leaving the system during a given interval of time may be called the universe of the stream for that period. For convenience, let us imagine that the universe is subdivided into individual parts whose measured characteristics will fall within a distribution pattern. We can then pick individual parts from this universe at random and expect them to conform to the principles of probability distribution.

The statistical question is, "How many parts shall we imagine exists in this universe, and how many of these parts must be removed to give a truly representative sample?" In practice the problem can be resolved by trial and error. A trial sampling frequency is selected and the samples are taken as scheduled. A statistical analysis of the data will establish the mean and the accepted variation limits. If the

sampling plan does not appear to be sensitive enough to detect significant effects of the variables the sampling frequency can be stepped up until significant information is uncovered.

To facilitate this technique the measurement of each part should be duplicated so that the buffering effect of measurement variability can be isolated from the in-process variations of the stream and the random sampling error.

Statistical Control of continuous chemical processes can be carried out, therefore, through a study of the relative differences occurring between successive precise measurements. Even though the measurement is biased, but is sufficiently precise to detect relative differences, it still can be used to advantage in the continuous control analysis because the analysis is made from related variabilities. The control depends on holding the variability within defined limits and decisions are made from correlation and significance analyses of the data.

The statistical control of continuous chemical processes can:

A. Give good overall process control under routine operating conditions.
B. Determine the effects of variables in the stream or on other variables.
C. Give correlations from which predictions may be made.
D. Give information which might lead to increased process efficiency or process simplification.

A successful example of this approach appears further along in the presentation.

With the background for the statistical control of continuous processes in mind, let us proceed with the steps leading up to the execution of the final plan.

The objectives of a control plan are the reasons for its very existence. Therefore, before selecting objectives, we should consider the limitations imposed by the equipment, the process, and the people operating the process. The limitations may not appear serious, but to save time and effort they should be scrutinized closely and the objectives proposed accordingly. Another consideration in the selection, of course, will be the availability of funds to carry on the project.

Having selected the objectives, a statement of the plan follows. In formulating this plan it is best that it tie in with existing controls and conform as closely as possible with the framework of present operations.

The plan must then be sold. It must be sold tc responsible superiors holding the purse strings, to the foreman directly supervising the process and to the operators who are manipulating the process but who may never have heard of statistical analysis. Perhaps others must also be convinced that the plan is feasible and that it has a very practical place in the scheme of the operation. The plan must usually be presented with tact and forethought, because it should have full acceptance by everyone concerned before further steps are taken. Many

well founded plans have fallen by the wayside before they were well underway through mistrust, disinterest and lack of cooperation. Teamwork is a must in the successful operation of a statistical control plan on a continuous process.

The preliminary phase, or trial period, starts with the first measurement. This period is full of uncertainties and speculation. Procedures are checked closely and the accuracy and precision of measurements verified. The methods of sampling and measuring, that were specified, must be scrutinized to detect flaws and make improvements.

As the data are being collected they are gone over very throughly. The statistical machinery is then set up for detecting sampling error and measuring process variability. Suitable control charts are prepared to chart in-process fluctuations on a continuous basis and a running record prepared of known changes which might affect the process. The analyses of the preliminary data should indicate inadequate sampling, if present, and a need for re-organizing either part or all of the original plan.

The trial period should also be used to acquaint everyone with the procedures that will be followed and to rehearse for the execution of the final plan. Special instructions in sampling (measuring) standards are reviewed with emphasis on taking samples which are representative of the material being measured at the time. The operations of the plan are coordinated and any discrepancies are smoothed out or eliminated so as not to plague the operations at a later date.

When the revised plan has been completed and approved, the project is officially ready to get underway. A starting time is scheduled, all final arrangements made, everyone alerted, and the plan is put into operation. As significant events show up, lines of action are determined and decisions made to direct the plan toward the chosen objectives.

To summarize, the recommended steps taken from the inception to routine operation of a continuous process control plan are given below:

1. Investigate all aspects of the process.
2. Prepare a background for the application of the statistical control technique.
3. Select and state the objectives.
4. Formulate the statistical process control plan.
5. Organize and sell the plan.
6. Progress through the preliminary phase.
7. Revise the plan and tighten the organization.
8. Execute the final plan.
9. Strive to attain the objectives.

An example of the approach of the method outlined is presented below.

I. Continuous Process Description:

A stream rich in the material to be chemically extracted enters the process in solution and is chemically treated, centrifuged and separated in four continuous stages. The fourth and final stage is a batch-type operation, as the process is not set up for continuous delivery of the solid product. The material to be recovered is at a relatively weak concentration originally. The concentration per milliliter increases at each of the four progressive stages of the process. Consequently, four spent side streams emerge as volume reduction occurs in the successive steps. The lower the units of material in the side streams, the better the process efficiency.

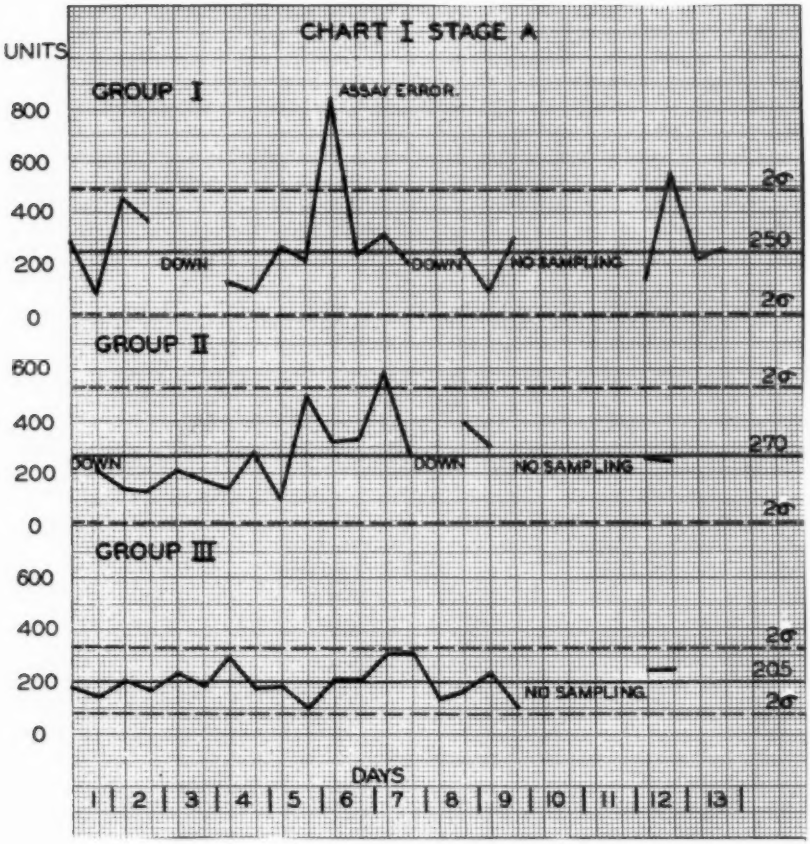
II. Objective:

The measure of success in the operation is a high percentage recovery of the incoming material of a predefined quality. The statistician's contribution to this objective consists in providing information and methods of analysis which lead to a higher percent recovery of acceptable material with less variability.

A. Trial Period

It was decided for the trial period that two samples would be assayed by each of the three shifts from the spent streams at Stage A, Stage B and Stage C. At the final stage of the process (Stage D), one sample was to be taken from each accumulated batch. As there is no schedule for batch completion, the sampling factor will be randomized at this stage. Unlike the previous three stages, the potency determinations are run by individual analysts. Consequently, the factor of analysts will be randomized also. The trial period was to cover approximately two weeks during which a group of workers remains on the same shift. This procedure confounded shift personnel (Group I, Group II, Group III) and the process behavior at each of the three time periods (8-4 period, 4-12 period, 12-8 period). However, if no significant differences appeared between groups (and/or process period), a longer trial period to separate the variability due each source would be unnecessary.

1. Chart I shows the results of the sampling at Stage A of the process.



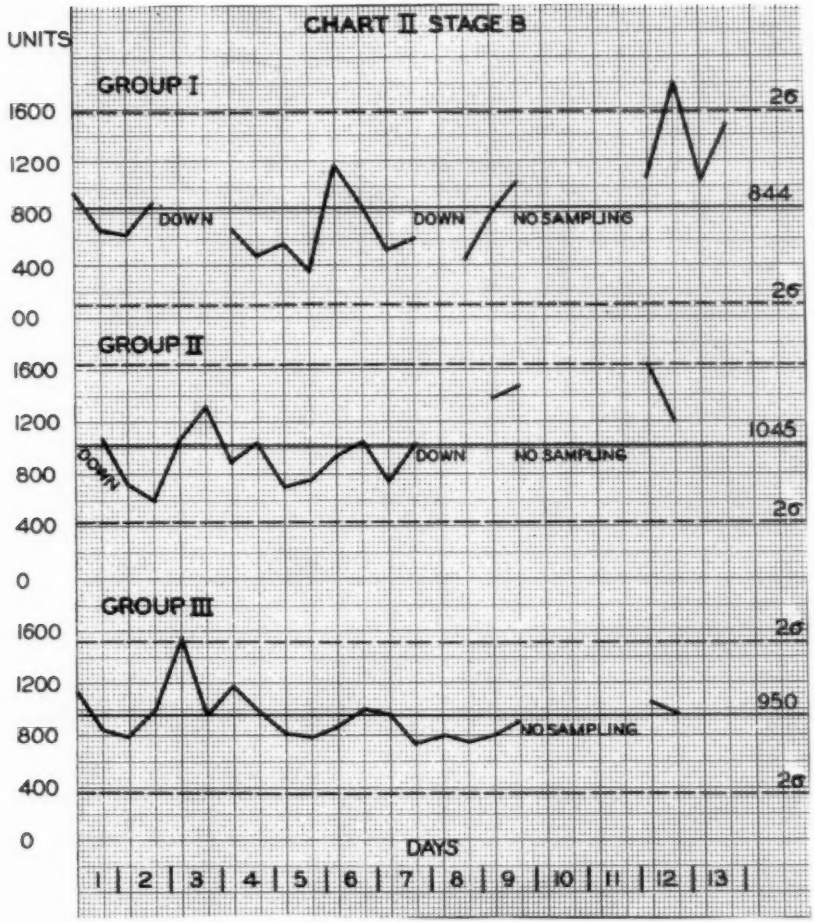
Group III has the lowest average units and also the least variability. This could be assignable to Group III, or it could be assignable to the process at stage A during the 8-hour period they worked. Table I gives the summary of variance for the 3-factor analysis of variance for Stage A during the trial run. The three factors are: days, groups and/or time periods and samples.

SUMMARY OF VARIANCE-STAGE A
TABLE I (ALL DATA CODED)

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARES
DAYS	2,063	8	258
GROUPS	1,145	2	572
SAMPLE	384	1	384
D x G	3,271	16	204
D x S	2,002	8	250
G x S	202	2	101
RESIDUAL	3,384	16	212
TOTAL	12,451	53	

When tested against a pooled residual, groups and/or time periods are at significantly different levels at less than the 10% level of probability. If the assay variability, which is known to be large, were removed from the residual, this difference might approach the 5% level of significance. Until the three groups have worked each of the three time shifts, this difference is confounded. It must be added that some of the missing values in this analysis were filled in by the usual method. This, therefore, somewhat limits the conclusions.

2. Chart II shows the results of the sampling at Stage B of the process.



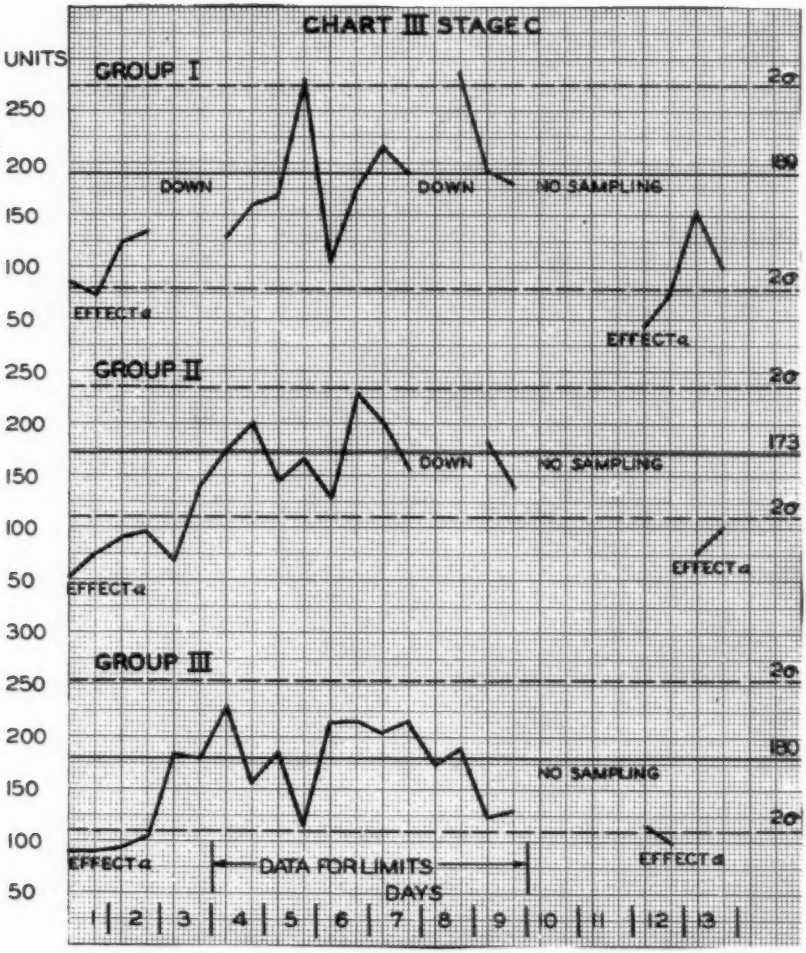
Group I has the lowest average units, although Group III again has the least variability. Table II gives the summary of variance for the 3-factor analysis of variance for Stage B during the trial run. The three factors remain: days, groups and/or time periods and samples.

SUMMARY OF VARIANCE–STAGE B
TABLE II (ALL DATA CODED)

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARES
DAYS	15,638	8	1,955
GROUPS	3,134	2	1,567
SAMPLES	1,661	1	1,661
D x G	15,201	16	950
D x S	3,420	8	428
G x S	646	2	323
RESIDUAL	7,625	16	477
TOTAL	47,325	53	

When tested against a pooled residual, the source days is significant at the .05 level of probability and approaches the .01 level. Groups and/or time periods are at significantly different levels at less than the 10% level of probability and the possibilities mentioned at Stage A also exist at this Stage. As with Stage A, missing values were calculated and consequently the exact level of probability is somewhat questionable.

3. Chart III shows the results of the sampling at Stage C of the process.



Group II has the lowest average units and also the least variability. This could be assignable to Group II, or it could be assignable to the process at Stage C during the eight-hour period they worked. It should be noted that the limits for the averages were not computed across all the data.

Effect a, as indicated on the chart, is the result of a routine process procedure at this stage. The objective at this specific stage is to decrease the variability around an average lower than the indicated plateaus. Table III gives the

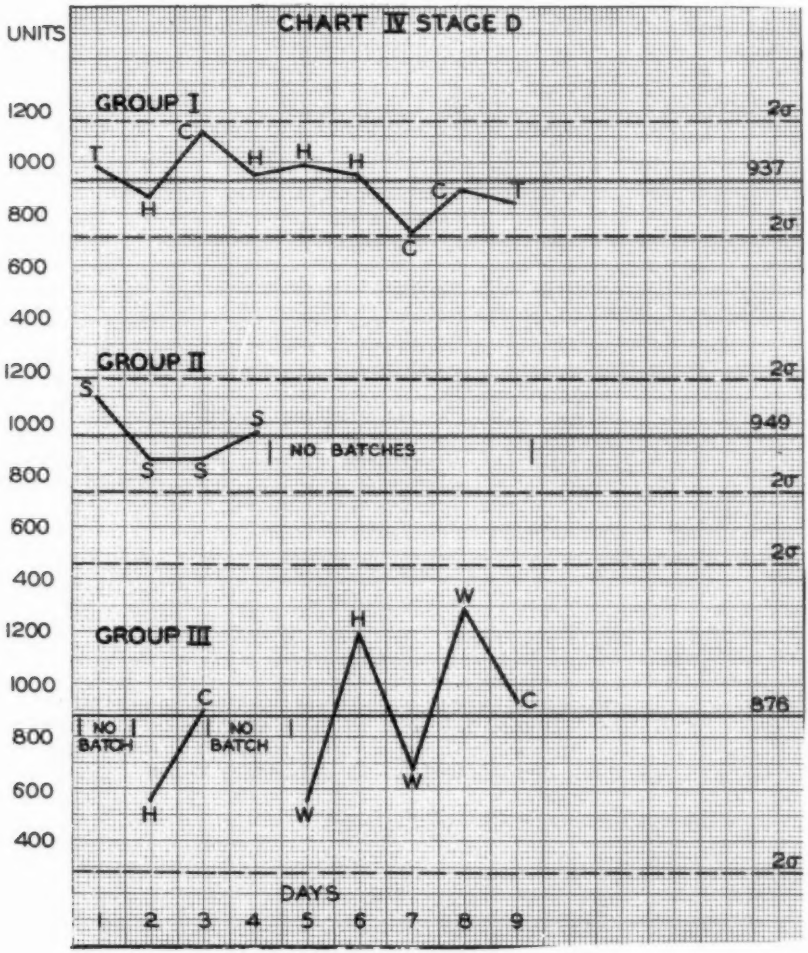
summary of variance for the 3-factor analysis of variance for Stage C during the trial run. The three factors remain: days, groups and/or time periods and samples.

SUMMARY OF VARIANCE-STAGE C
TABLE III (ALL DATA CODED)

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARES
DAYS	95,338	8	11,917 ***
GROUPS	1,438	2	719
SAMPLES	3,008	1	3,008
D x G	16,447	16	1,028
D x S	10,575	8	1,322
G x S	9,081	2	2540
RESIDUAL	35,296	16	2206
TOTAL	171,183	53	

When tested against a pooled residual, the source of variance days is significant at the .001 level of probability. As with Stages A and B, missing values were calculated and consequently the exact level of probability is somewhat questionable.

4. Chart IV shows the results of the sampling at Stage D of the process.



Group III is at a lower average level and has a much greater variability pattern. This could be caused by the process at the 8-hour time periods, Group III operation, or in this instance, by the chemical analysts who ran the determinations (identified by letters). According to the F-test, Group III has a significantly greater variability than Group I at the .02 level of probability. (There are insufficient degrees of freedom when compared to Group II). The data at this stage, as previously indicated, are unbalanced so an

analysis of variance was not used.

Conclusions and Further Action Recommended From the Trial Period

a. There was evidence of group differences. As previously explained, this could arise from:

1) Sampling differences
2) Assay analyst differences
3) Processing differences

Consequently, a program to standardize further sampling procedures was established. The analytical laboratory was directed to run a designed experiment to ascertain between analyst differences and within analyst variability as shown by replication. Finally, as the analyses of variance showed no significant difference between the samples taken on the same shift, the first routine control plan called for one sample per shift to continue for a twelve-week period, so as to segregate group effect from process effect with between-two-week periods as the residual.

b. The significant between-day difference found at Stages B and C was expected. Process effect in Stage C is present in Stage B also, although to a lesser degree. This lack of independence is common in a continuous process.

The question of pooling or not pooling in analysis of variance is open. In these analyses of variance pooling was used, although the decisions would not have been altered if this had not been done.

B. Results of Routine Control Plan:

1. Stage A

The factor, groups and/or process time period, which was originally considered as possibly real, disappeared with standardization of sampling technique. The analysis of the data from the trial period showed the departmental supervision where to investigate. It was the process foreman who actually corrected the difficulty. This stage of the process is very stable and the lack of a group X shift interaction indicates each group performs similarly on each shift.

2. Stage B

The difference between groups is significant at the .001 level of probability. Group I has significantly lower spent stream potencies which indicates

either better process operation or sampling technique. The latter is unlikely with the intensive investigation just completed. The group X shift interaction arises from process effect a which does not follow a standard pattern. The process at this stage is also very stable when the above variabilities are segregated.

3. Stage C

There is no evidence of any difference between groups or shifts. The group X shift interaction arises from process effect a which actually takes place at this stage and, as previously mentioned, is felt somewhat at the prior Stage B.

4. Stage D

The analysis of variance shows nothing significant, It is of interest to note process effect a has no direct influence at this stage. The reason for this, as will be seen presently, is not because Stage C and Stage D are unrelated, but the analysis of variance does not permit its detection. The chemical analyst differences had been satisfactorily eliminated and the determinations at this stage are now more precise.

Table IV is the summary of the results during this phase of the problem.

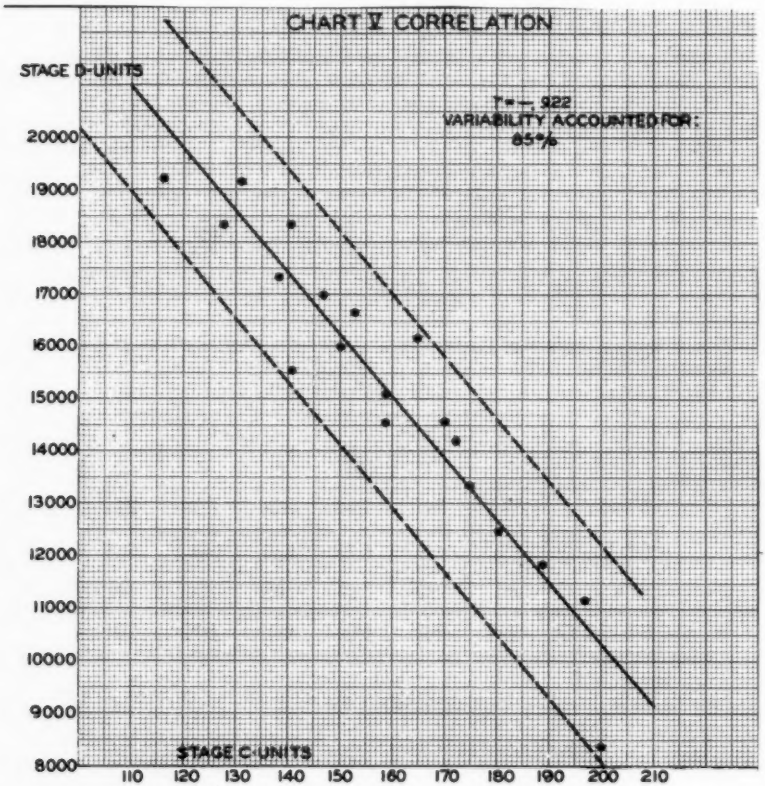
SUMMARY OF AVERAGES AFTER 12 WEEK PERIOD
TABLE IV

	GROUP			SHIFT		
	I	II	III	8	4	12
STAGE A	234	246	214	238	234	222
STAGE B	456	753	875	659	814	812
STAGE C	141	139	139	140	139	141
STAGE D	684	687	701	680	670	701

It will be noted that the last three stages are at lower levels than recorded during the trial period. Process changes had been made in Extraction and conditions had been altered in the production of the rich stream which are beneficial. Extreme care must be exercised to ascertain the effects of changes. Lengthy studies depending upon process "sameness" must be avoided. Continuous statistical techniques must be employed for continuous type processing.

C. Simultaneous Process Investigation:

Every process has its characteristics which can be measured and studied. Temperature, pressure and relative acidity are examples of a few. Every factor which may effect the incoming rich stream must be taken into account as well as sources of variability within the continuous process. The best available method for statistical analysis is the correlation technique. Chart V shows one of the numerous correlations studied in this specific problem.

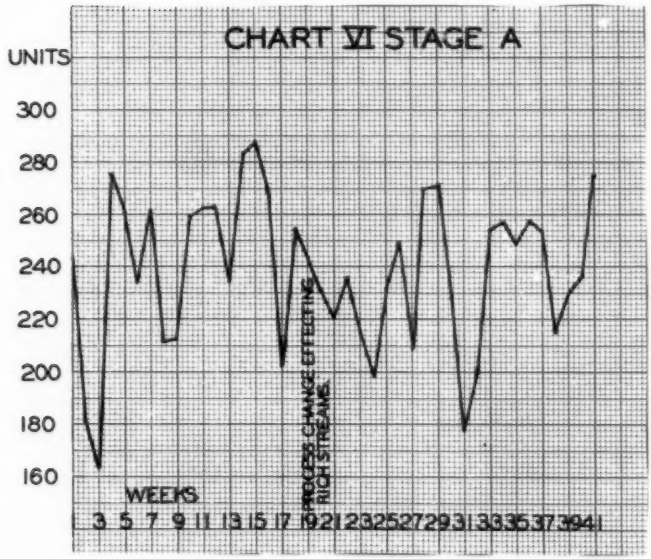


The average weekly units of the spent side stream at Stage C are on the X-axis, while the average weekly units of the spent side stream at Stage D are on the Y-axis. As indicated, the relationship is highly significant with a correlation coefficient of -.922. Until this detrimental relationship was broken by a process change, there was no possibility of increasing the efficiency of the process at these stages and consequently increasing the percent recovery beyond a definite limit.

D. Histories of Each Stage Showing Effects of Investigation

1. Stage A

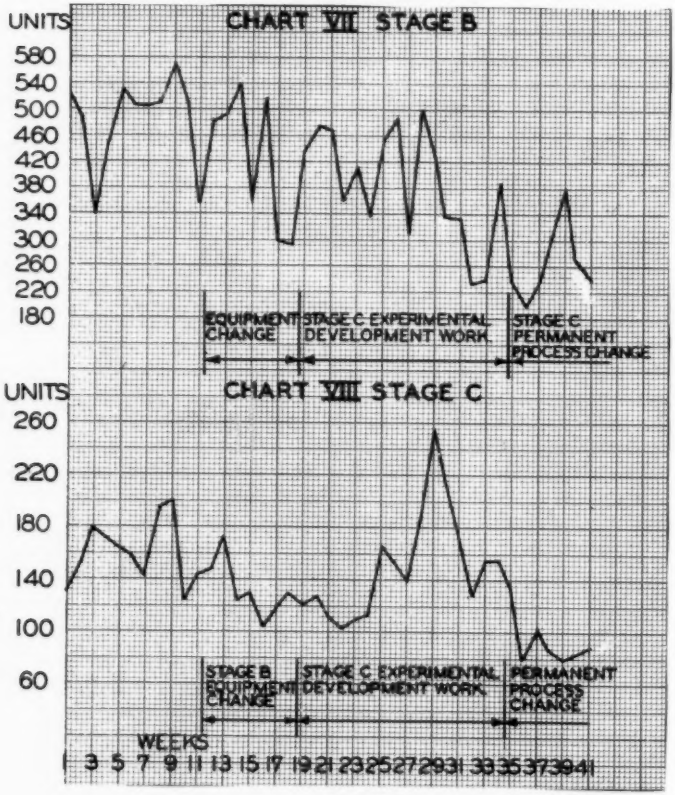
Chart VI is a plot of the average units of the spent stream for forty-one consecutive weeks.



As can be seen, there is much unassigned variability at this time. However, as this is the point of least weight, more effort has been concentrated on the subsequent stages.

2. Stages B and C

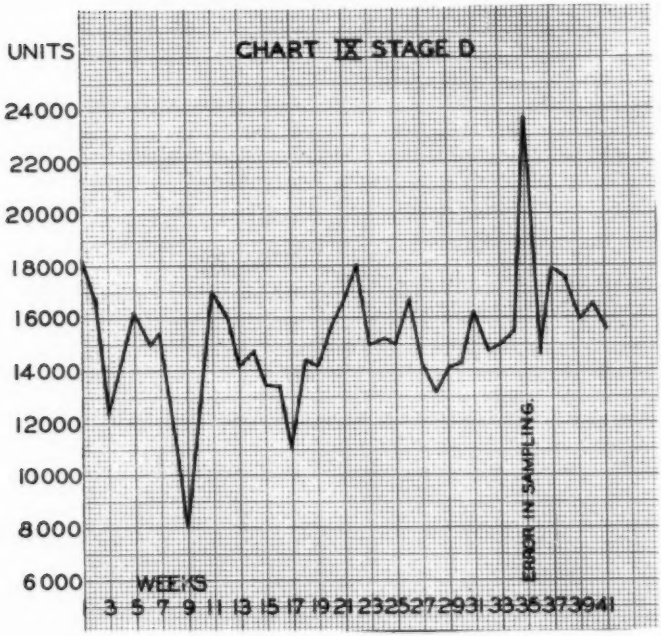
Charts VII and VIII are plots of the average units of the spent streams for the same forty-one consecutive weeks. These are presented together because they are inter-related.



An equipment change at Stage B was introduced during the twelfth week. As is generally the case, a period of time was required before the effect could be detected. After a lag, the effect was very definitely noticeable at Stage C. Development work on Stage C caused an increase in the spent stream potency at Stage C. This work, however, caused a continued downward trend at Stage B. A major process change at Stage C during the thirty-fifth week reduced the variability of the process at Stage C at a lowered average level. The level and variability at Stage B were uneffected.

3. Stage D

Chart IX is a plot of the average units of the spent stream for the same time period.



It will be noted that the average units are much higher than those shown previously in Table IV. This is a result of continued investigation of the chemical assay prompted by the findings of the trial period. This increase in assay accuracy accounted for some of the process loss previously unassigned. However, of greater importance from the point of view of this paper, is the further evidence, that regardless of bias in measurement, a process can be intelligently studied and improved as long as the measurements are precise. The chart shows Stage D remained stable and unaffected throughout this entire period.

The example emphasizes the application of statistical procedures in a typical continuous chemical process and in accordance with the philosophy outlined. The objectives were defined, the control plan tentatively outlined, the trial period evaluated, action taken, and the final plan was put into effect.

Many of the data uncovered were not entirely understood, however, the process was placed under control and valuable information was obtained during the study. The control plan should continue and the objectives of higher process efficiency, a greater yield of acceptable product, and a reduction of processing costs realized.

References:

(1) Kirkbride, C. G., "Chemical Engineering Fundamentals"; Mc-Graw-Hill Book Company, N. Y., N. Y., pp 58, 1947.

(2) Brown, C. G., "Unit Operations"; John Wiley & Sons, Inc., N. Y., N. Y., pp 1, 1950.

(3) Mueller, R. K., "Effective Management Through Probability Control"; Funk-Wagnals Company, pp 3, N. Y., N. Y., 1950.

(4) Shewhart, W. A., "Economic Control of Quality of Manufactured Product"; D. Van Nostrand Company, N. Y., N. Y., pp 6, 1931.

(5) Ibid, pp 53.

SOME ASPECTS OF EXPERIMENTAL DESIGN

Lee Crump
University of Rochester
Atomic Energy Project

In the time available here it is not possible to discuss in a very useful fashion more than a limited area of the broad problem of planning and designing experiments. I shall attempt to discuss in a general way certain aspects of the problem that are of particular interest to the statistician. My plan is to consider in some detail a simple illustrative situation, raising in connection with it certain questions that seem to be representative of those which arise in the planning of most experiments.

Suppose that we are considering the purchase of a machine or machines to perform some manufacturing operation, and that there are two producers of machines designed for the purpose at hand. Our broad problem is to decide which producer to patronize. We shall assume that a point has been reached where the decision hinges on the answer to a question of the following form:

Which producer's machines make the values of x larger?

The variable x is some continuous quality characteristic of items turned out by the machines. It is proposed to answer this question by experimentation, and we have now to consider the problem of planning a "good" experiment for the purpose.

The first point to be noticed is that the question as stated is not sufficiently specific. This is generally the case. The question which originally suggests experimentation is usually one which is too general or imprecise to be answered experimentally. The first problem, and the one with which most of this discussion is concerned, is that of stating a question which can be answered experimentally and which at the same time is worth answering.

A brief consideration of the situation will indicate the lines along which the question has to be modified. We probably know that the values of x in a series of items produced by one of these machines vary appreciably among themselves, and that a similar series from another machine of the other (or even the same) producer would yield x-values which also varied and overlapped, to some extent, those in the first series. If we did not know or at least suspect that this were true we should probably not be considering an experiment. The word "larger" in the original question must mean "larger in some average sense". Our interpretation of this is that in some way as yet unspecified there are two frequency distributions (one for each producer) involved and what we really want to know is which of these two distributions has the larger mean. The original question implies, as does our discussion up to this point, that the variabilities in the two frequency distributions are not pertinent to the decision. We shall not concern ourselves with just how this viewpoint might have come about. It will be necessary for us to consider these variabilities below, but not as a part of the question we are attempting to formulate.

We must now turn our attention to a more precise description of the frequency distributions which we have mentioned. Several preliminary remarks are in order:

1. The frequency distributions about which it is useful to know will in this case and in general be such that they can only be investigated partially, i. e. on a sampling basis.

2. In the present state of our knowledge we only know how to "answer" questions about frequency distributions on the basis of samples from them when the samples are random. (N. B. In this discussion a question about a frequency distribution is "answered" only if some precise statement of the probability with which the answer is correct can be made also.)

3. Hence our experiment must be such that the data it provides constitute random samples from frequency distributions which are pertinent.

Most of the statistician's problems in connection with planning experiments result from this state of affairs. For example, the frequency distributions of interest in this case are presumably those which would be generated by letting each of the two types of machine turn out a large number of items under the conditions in which they are to be used, determining x for each item, and arranging the values obtained into frequency distributions. In order to "answer" questions about these frequency distributions our experiment must represent a random sample of whatever we define to be "the conditions in which they are to be used". Such a requirement can almost never be completely satisfied in practice. The experiment for example has to be conducted in a relatively short time interval and it is difficult to see how any reasonable time interval can be made to include an anywise random sample of the conditions of ordinary routine operation. Perhaps the experiment is planned for the month of January. It is likely that many of the operating conditions are different in January then in July, say. We may feel on the basis of our experience and knowledge that the conditions which differ between the two months are unrelated to the behavior of the machines, and that conclusions about January are applicable to July. This feeling may be quite correct, but the experiment in January neither defends nor refutes it. It must stand on extra-experimental evidence. In nearly all experiments there will be an inferential step between the question which the experiment can "answer" and the question which we would like to "answer" which cannot be made rigorous in a statistical sense. In practice we can only attempt to make this step as small as practicable.

Let us assume now that the necessary decisions have been made about the conditions under which the experiment is going to be conducted, and turn our attention more particularly to the actual experimental material. Consider first the machines to be used. We can illustrate the points we have in mind here most easily by distinguishing two cases within our illustrative situation.

In the one case assume that only one machine is to be purchased and that we are not likely to have to replace it for a long time.

Assume further that the machine purchased will be one of two available for experimentation, one from each producer. In the other we suppose that a relatively large number of machines will be purchased without further experimentation from one producer picked on the basis of this experiment. We assume also in this case that just one machine from each producer is available for test.

The difference between the two cases implies a corresponding difference between the frequency distributions which are of primary interest. There are first the two distributions that would be generated by the two machines at hand under the conditions of the experiment. Let A denote the mean of the one of these distributions generated by the first producer's machine and B the mean of the other. In the first case our question refers to A and B. Now we can think of distributions similar to the two just mentioned which would be generated by other machines supplied by each of the two producers. For each producer there is a frequency distribution of these machine means which itself has a mean. Denote these means of machine means by A' and B' for the first and second producer respectively. In the second case our question refers to A' and B'. Any information which the experiment furnishes about A and B is useful in this case only insofar as A and B can provide information about A' and B'.

We can "answer" questions about A and B experimentally no matter how the two machines at hand were selected from the total production of the two producers. In order to "answer " questions about A' and B' the two quantities A and B must be random samples from the distributions defining A' and B'. The two machines used in the experiment must represent a random selection of the machines made by each of the two producers. (There will still in general be some difficulty about "answering" questions concerning A' and B' unless we know or can estimate the variability in the distributions defining them. This difficulty is irrelevant to point at issue here.) In summary, the importance of randomness in the selection of the machines for test depends entirely upon the situation. Many experiments fail because a mistake similar to that of testing a demonstration (non-random) model from each producer in order to "answer a question about A' and B' is made.

For the balance of this discussion let us assume that we are in the first situation, i. e. interested only in the values of A and B. An experiment involving the two machines at hand is pertinent and a decision about the general conditions under which the experiment is to be conducted has been made. There remains to be decided just what experimental units are to be used and how they will be distributed to the two machines.

For simplicity let us assume that the raw material for the operation in question enters the machine in the form of a one-foot bar. Suppose further that it is received in the form of bars twice this long and that the immediately preceding operation is just that of cutting the two-foot lengths in half. In a broad way the experiment will consist of feeding a series of one-foot bars into each machine and the experimental data will be measured values of x on the items produced. From the data we will "answer" some questions about the relative sizes of the mean values of x in the two distributions of which the experimental data from the first and second machine are a random sample.

Now in "answering" any question about these means we know that on the whole the probability of obtaining a correct answer is a decreasing function of the variability in the distributions defining them. The magnitude of this variability can be influenced to some extent at least by the bars which we admit in the experiment and we have some freedom of choice in this regard. We must at the same time however remember that we want our conclusions to apply essentially to all of the bars which will be used in routine operation. The desire to "answer" with a high probability of being correct and at the same time the desire to have the answer apply under wide conditions are not entirely compatable. We have to reach some reasonable compromise.

There are several obvious ways in which we might select bars to use in the experiment and a brief consideration of each of them will be instructive. The simplest procedure would be to just take the desired number of bars out of whatever is the normal stream of supply to machines like the ones under test. If the series of outgoing bars from the preceding cutting operation is in statistical control, then by definition we have a random sample of the normal supply of bars and our experimental conclusions may apply to the whole of the normal supply. If this series is not in control we will be unable to reason with complete statistical validity to the whole of the normal supply. In the latter case we may feel quite safe in assuming that the essential parts of our experimental conclusions apply to the whole supply, but once more this assumption must stand on extra-experimental evidence.

In either case if the inherent variability from one bar to another is large we may require an excessively large number of bars in the experiment to reach some desired probability of "answering" the question asked correctly. We might undertake to pick out of the normal supply of bars a group for experimentation which was as homogeneous as possible with respect to certain features which were felt to be related to the final value of x. In this way we might be able to raise the probability of being correct for a given size of experiment, but the question "answered" by the experiment would doubtless be even further removed from actual routine operation.

There can be no categorical answer to the question of which of the alternatives so far suggested is the better. Each particular situation has its own special properties and the decision has to be made more or less independently in each case. If we do not have a series of bars in statistical control it is difficult to see how we can ever obtain an even approximately random sample of the whole of normal supply, and in such a case we will always be forced to "answer" a question which does not refer (without extra-experimental assumptions) to the exact conditions in which the answer will be applied. In many cases the higher probability of being correct by utilizing a homogeneous non-random selection of the potential experimental material may outweigh the advantages of an entirely random selection even though such could be made.

If our experimental material is not a random sample from some meaningful distribution as in the case of the selection of especially homogenious bars, it is not obvious just what distribution if any we have sampled randomly. One of the primary functions of objective randomization in the assignment of experimental units (in our case bars) to the treatments (in our case machines) is to assure that the experimental observations do constitute random samples from meaningful

distributions. We do not choose to discuss this aspect any further here but refer the reader to Design and Analysis of Experiments[1] by O. Kempthorne.

Since the raw material originally is received in two-foot bars a third possibility for selecting the experimental bars suggests itself. It is quite likely that two bars cut from the same piece are more homogeneous than two bars selected without regard to their origin. We feel instinctively that if in our experiment we allocate the bars in pairs, one to each machine, this homogeneity of bars in the same pair ought to improve the experiment. Hence we may select instead of individual bars, pairs of bars cut from the same piece, allocating one bar of each pair to each machine. In adopting this plan we shift our attention from two frequency distributions, one for each machine to a single frequency distribution of differences between the x-values, each difference being the x-value for a bar in the first machine minus the x-value for its partner in the second machine. Questions about the means of the two distributions maybe reformulated into equivalent questions about the mean of the single distribution. The probability of "answering" questions about this mean correctly may be higher than in the experiments previously proposed.

Such an experiment is called a Randomized Block experiment. It is the simplest of the formal experimental designs which attempt to take advantage of internal homogeneities of the experimental material or conditions. The statistical analysis of the data from such an experiment is based on assumptions about the frequency distributions involved which are somewhat more restrictive than those required in the previously mentioned experiments. For a detailed discussion of these matters see the book by Kempthorne previously referred to and Experimental Designs[2] by W. G. Cochran and G. M. Cox. I should like to point out here that there is a regrettable tendency to overlook the more restrictive nature of the assumptions that are required to validate the analysis of data from more complicated experimental designs.

I have attempted in these remarks to indicate in a broad fashion some of the factors that enter into recognizing clearly just what questions can be "answered" by a proposed experiment. I have intended to emphasize the fact that these questions are usually severely limited and that we must often rely heavily upon extra-experimental evidence to give our experimental conclusions broad applicability.

1. John Wiley and Sons, Inc., New York 1952
2. John Wiley and Sons, Inc., New Yrok 1950

PRACTICAL APPLICATION OF STATISTICAL QUALITY CONTROLS IN THE AIRFRAME INDUSTRY

Phil Harr
Fairchild Aircraft Division

INTRODUCTION

Whenever a new method or process has been developed in any field, it has been customary for claims of a somewhat exaggerated nature to be made. Too often the similarity of the new method to methods already in use is not considered and too often in our enthusiasm over the new tool, we tend to think of it in terms of a cure-all.

With the test of time, these new methods have always assumed their proper perspective and have become recognized for their worth as one of the tools with which we may work and are utilized only where it is logical to do so.

For us, Statistical Quality Control is no cure-all. It is a good tool and one which is improving with use. It is the purpose of this presentation to explain the current use and limitations of Statistical Quality Control techniques in airframe production, as practiced by the Fairchild Aircraft Division, in the hope that our problems and solutions may contribute to the practical application of such techniques in industry as a whole.

Statistical Controls are useful tools in the solution of many production and Quality problems. However, the stress which must necessarily be placed upon quality in airframe manufacture imposes a practicable limit upon their application in this field. The extent to which we may efficiently utilize these statistical tools depends on a careful balancing of the cost against the resultant savings or the improved quality effected.

THE OVERALL PROBLEM

Examination of the Flow Chart, Fig. (1), will show that items which are procured outside of our plant are divided for Quality Control purposes into the following categories:

1. Raw Material in its various forms - the major ones are as listed.

2. Hardware - These are standard small items such as bolts, nuts, rivets, etc.

3. Equipment - These are in general functioning assemblies either purchased direct from the manufacturer or in case of engines, propellers, wheels, radio, instruments and similar items, they may be furnished by the Government.

4. Sub-contract Items - These are parts or assemblies designed by Fairchild and assigned to our sub-contractors for manufacture. Such items range from machined castings to complete assemblies, i.e., rudders, ailerons and flaps.

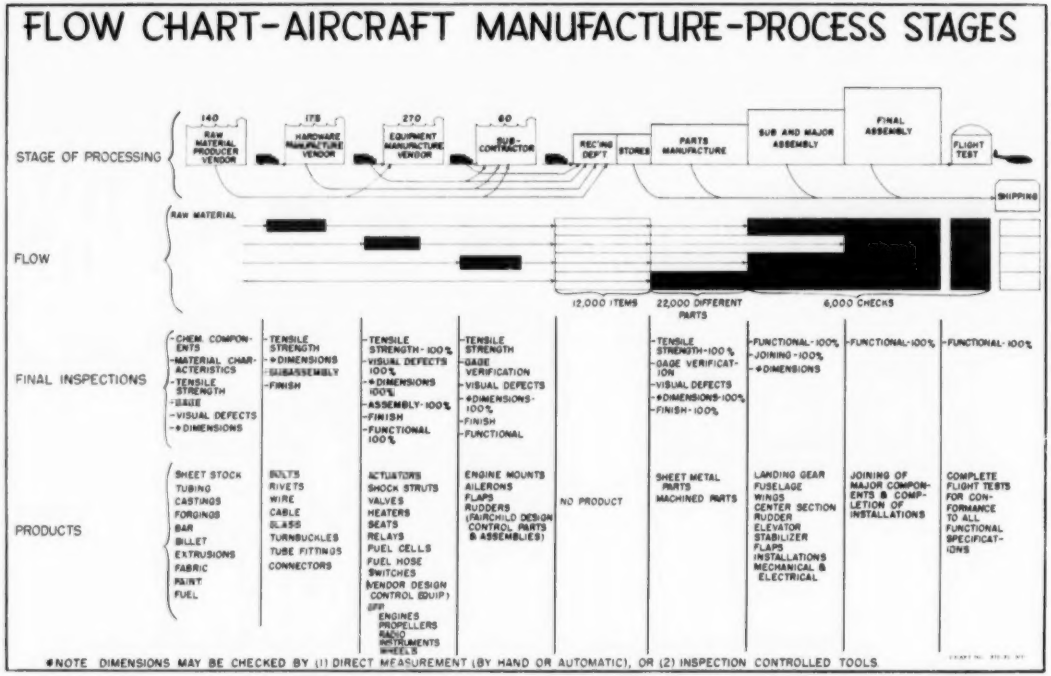
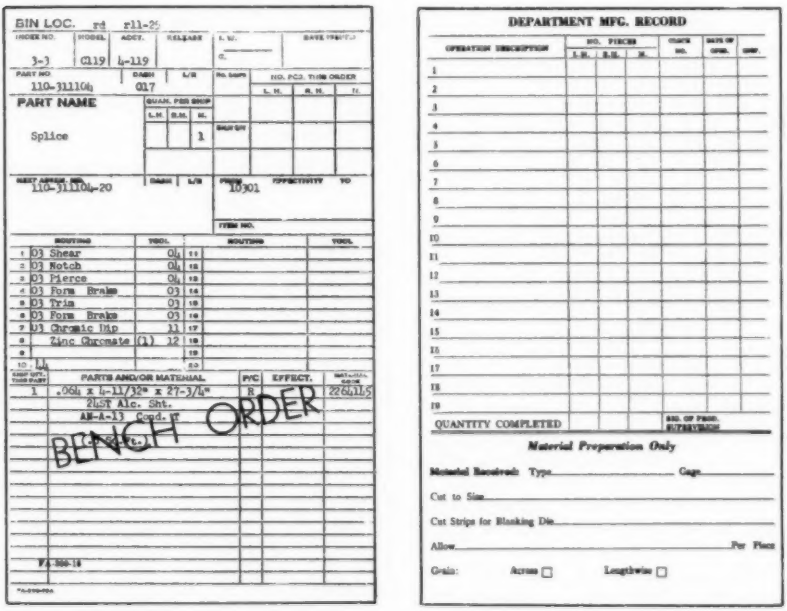


Fig. (1) Flow Chart - Aircraft Manufacturing & Process Stages Blocked in spaces indicate points in flow where manufacturing operations take place producing new or changed characteristics

A typical airframe model contains over 12,000 different items of raw material, customer furnished and purchased parts. Each of these items has from several to hundreds of characteristics, many of which are impossible or impractical to inspect after receipt at our plant.

These same materials and parts are processed through manufacture by over 22,000 work orders, each having from four to six or more operations requiring inspection. Each of these operations has from several to many characteristics. It might be well to inject here what we mean by Work Orders. The front and back of a typical Work Order form for a small sheet metal part is illustrated in Fig. (2). These are tickets issued by Production Planning which call for a given quantity of a particular part or assembly number on a particular schedule. Routing information is given by major departments and space is provided for indication of progress by Production and Inspection stamps. These tickets refer to detailed Operation Sheets which are held in the Departments and which spell out in more detail the work to be done and the tooling to be used. A Work Order routing for example, may refer to operations producing several dimensions on the part.


BIN LOC. r4 r11-26

INDEX NO. 3-3 MODEL C119 ADDT. 4-119 RELEASE

PART NO. 110-311104 DASH 017 L/B

PART NAME Splice

QUAN. PER SHIP L.H. 1

NEXT ASSY. NO. 110-311104-20

	ROUTING	TOOL		ROUTING	TOOL
1	01 Shear	04	11		
2	01 Notch	04	12		
3	01 Pierce	04	13		
4	01 Form Brake	03	14		
5	01 Trim	03	15		
6	03 Form Brakes	03	16		
7	03 Chromic Dip	11	17		
8	Zinc Chromate (1)	12	18		
9			19		
10			20		

	PARTS AND/OR MATERIAL	PYC	EFFECT.	MATERIAL CODE
1	.064 x 4-11/32" x 27-3/4" 2ST Alc. Sht. AN-A-13 Cond. F	R		2264115

BENCH ORDER

DEPARTMENT MFG. RECORD

OPERATION DESCRIPTION	NO. PIECES L.H.	R.H.	M.	CHECK NO.	DATE OF OPER.	INSP.
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
QUANTITY COMPLETED				NO. OF PROD. RUN/STATION		

Material Preparation Only

Material Received: Type Gage

Cut to Size

Cut Strips for Blanking Die

Allow Per Piece

Grain: Across ☐ Lengthwise ☐

Fig. (2) Production Work Order

Assembly is controlled by an additional 6,000 inspection sign-off points and here again, each sign-off involves many characteristics which must be given careful consideration by inspection. A typical sign-off book for an elevator assembly contains several pages of items which describe stages in its production progress and inspection points to be checked.

Let's select one typical item - for example, sequence item #1-18 which calls for riveting, Fig. (3). This item reads "CHECK RIVETING OF CENTER BOXES BEFORE CLOSING. Ref. IMB-G-2;41 MPS-FA-14"

INSPECTION SIGN-OFF

DEPT. 109	MODEL C119	SPECIFICATION 10857- &Spares	ASSEMBLY NAME Frame Assem. - Elev.	ASSEMBLY NO. 110-220050-110
DATE ISSUED OR REVISED	wm 2-4-53	SHIP OR SERIAL NO.		PAGE OF 4

WORK STA. OR SEQUENCE NO.	TITLE	EFFECTIVITY FROM	EFFECTIVITY THRU	DATE	INSP.	F.A.	A.T.
Cont. 1-18	(C) Check riveting of center boxes before closing						
	Ref: IMB-G-2;41 MPS-FA-14						
1-19	Check assem. & erco riveting of rib and angles						
	Ref: IMB-G-2						
1-20	Check assem. & erco riveting of rib and angles						
	Ref: IMB-G-2						
1-21	(A) Check assem. location and drilling of parts in assem. jig & riveting where possible						
	(B) Check attach points in jig						
	(C) OK to remove from jig						
	Ref: IMB-G-2; 7; 41						
1-22	(A) Check complete riveting in clean-up						
	(B) Check install. of tab control mech. & spring tab cartridge mech.						
	(C) Check install. of vibration dampeners-to be depressed by rods .09 max to .06 min Ref: BAI 15 or 110-221050						

FA-SIO-27A

Fig. (3) Typical Inspection Sign-off Sheet

Let's follow this Ref. IMB-G-2 or Inspection Methods Bulletin on riveting practices through to determine just what it is that the inspector will look for in this case.

Among other things the inspector must check:

1. Proper type of rivets
2. Proper size
3. Proper quantity and pattern
4. Proper forming, Ref. ID-5

Let's stop here - what is a properly formed rivet? The inspector could pass unsound riveting or he could insist on picture book perfection and cause excessive and unnecessary costs. The inspector, however is

supplied with another reference, in this case, an Inspection Data Sheet, which illustrates typical conditions encountered in riveted joints together with acceptance limits.

It might appear that the many references are too detailed and time consuming, but please remember that literally thousands of rivets are used in every airplane and for the point of this discussion, the quality analysis of each of these rivets must be made on a consistent basis.

In actual practice, it is not as bad as it appears at first glance because the production worker and the inspector rapidly become completely familiar with these references and make consistent decisions without looking up the actual paper.

Needless to say, no control of any type may be achieved without first establishing standards which carefully define the quality required. These standards are in terms of characteristics which may range in one case from the hardness of 4130 steel to decibels of allowable radio noise from an actuator or, in another case, from the pressure output of a hydraulic pump to the surface roughness in micro inches of a machined finish. Only with standards which eliminate opinions can we achieve fair comparisons and only with comparisons can we evaluate performance and progress.

It is important for the sake of efficiency that each step in production progress be designed to insure the earliest possible detection of defects and that provision be made for prompt corrective action. It is a critical necessity in aircraft that no significant defect go uncorrected as a single improper characteristic in the product can easily take a heavy toll in lives and property.

Constant analysis of this progress must provide for modifying inspection controls in accordance with current results, thereby obtaining high quality work with reduced inspection and at the same time, insuring prompt correction of low quality work. It is in this analysis that we apply various types of statistical controls to suit the problem.

THE CONTROLS

It is of fundamental importance from a management viewpoint that controls be designed and applied first from an overall picture and, then in the interests of efficiency, focusing on details only where and to the extent necessary to keep the overall quality consistent with management and customer requirements.

Analysis of Fig. (1), will immediately reveal that quality control is not an in-plant problem alone. Many characteristics are created in the plants of our suppliers and any failure to control the quality at this early stage of manufacture is like building a house upon the sand. It is also apparent that quality control, like manufacturing processes, requires a job breakdown even in our own plant due to the thousands of characteristics created or altered to arrive at the final product. Authority for detail control must be set up at the floor level with provision for bringing only critical or unusual volume difficulties to higher levels in the organization.

Any attempt to approach our overall quality problems in terms of individual parts and assemblies results in an almost hopeless confusion

of paperwork. We submit that it is absolutely necessary to think in terms of characteristics for it is only when one speaks of a characteristic that the terms "final inspection", "100% inspection" or "sample inspection" begin to mean anything. Contrary to ones first impression, the consideration of quality characteristics immediately simplifies our problems for we find that of the multitude of parts almost all have many characteristics in common. This we believe is clearly illustrated in Fig. (1) under the horizontal column entitled Final Inspections.

Here we have, in relatively few subjects, described the inspection of thousands of materials, parts and assemblies. True, the actual tensile values, for example, have not been specified for our various materials and heat-treatments, but even when we expand this general subject to include the values used, we find they are relatively few in number as compared to our total parts. In other words, we make many different parts of identical tensile characteristics.

Expanding our thinking in terms of characteristics, we find upon analysis of our work orders, that large groups of parts have common characteristics. Each of these characteristics has an "end point" where it logically receives its ultimate or "Final Inspection". For example, the final point that the tensile strength of a sheet metal part may be tested is after fabrication and prior to assembly. There we can detect inadvertent mixing of material and improper processing and by proper identification insure that only parts of acceptable heat-treat characteristics are incorporated into the airplane. This point is then an inspection point as fundamental to the safety of the structure as the operation of the engine is to the flight characteristics. Therefore, only 100% inspection is allowable. For example; we at one time stated that we run 100% inspection on all actuators in our Receiving Department. This is only a part truth because actually these actuators contain literally hundreds of characteristics which it is impossible to inspect at this point. What we were really doing was inspecting the actuators 100% for their operational and visual characteristics only. The control of other characteristics is of necessity by other methods.

Let us illustrate. Fig. (4) shows a typical hydraulic actuator. This one is used to retract the main landing gears in the Fairchild C119 airplane. This actuator contains 29 individual parts, each of which have dimensional, strength and other characteristics which affect the operation of the assembly.

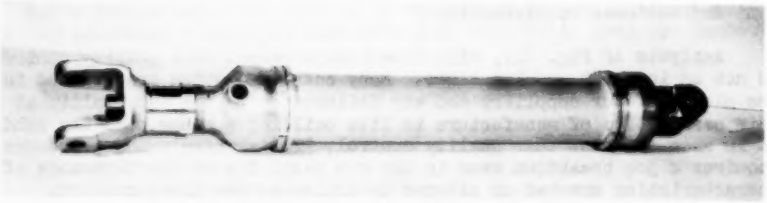


Fig. (4) Hydraulic Actuator, Landing Gear

How do we control these characteristics?

First, we insist that our engineers spell out exactly what is wanted in terms of performance, overall dimensions, finishes, and any other pertinent requirements.

Second, we survey the prospective vendors facilities and organizations to determine their ability to produce to our requirements.

Third, we completely dis-assemble and functionally test first production units received to determine if they meet all detailed requirements.

Fourth, we functionally test all production units either prior to or after installation in our airplane.

Fifth, reports from our Field Service Representatives are carefully analyzed for any indication of difficulty. Corrective action is prompt and complete and may involve additional teardowns, simulated service testing, re-designs and replacements as required.

SAMPLING INSPECTION

Up to now we have talked very little about Statistical Quality Controls because before we can clearly explain how we control we must arrive at a common understanding as to what it is we are trying to control. Now let us see how we endeavor to apply quality controls as efficiently as possible:

We were speaking of the heat-treat characteristics of sheet metal parts. We just stated that 100% inspection is necessary at this "Characteristic End Point". Where then, is there any room for sampling inspection?

1. At the Mill - Here we find that mill runs or lots manufactured under identical conditions assure that samples are truly representative of the lot.

2. At Receiving for Stock Control - Here we must insure against mixing or damage during warehousing and shipping operations and thereby insure that our stock rooms contain material which is for the most part, satisfactory for use in the shop.

3. In Process in our Shop - Here we can insure that mixing, latent defects, handling damage, or faulty workmanship are detected at an early stage and thereby minimize the amount of rework and reduce the cost of scrap to a bare minimum.

Previous to the "End Point" for tensile strength characteristics, the hardness of sheet metal parts is inspected almost solely to reduce rework and scrap or to insure substantially usable stock. Therefore, the percent of inspection may well be determined on the basis of previous results or, in other words, statistically. Similar analysis of items in other categories point the way to studies of inspection results and the adoption of acceptance sampling plans prior to the "characteristic

end point" as defined above.

It might be well to mention that where "characteristic end points" require 100% inspection, it is, of course, of first priority importance to develop efficient methods and the utilization of automatic checking equipment. An example of such equipment is shown in Fig. (5). This hardness certifier checks the heat treatment of sheet metal and signifies acceptability by automatically applying the heat treat stamp.

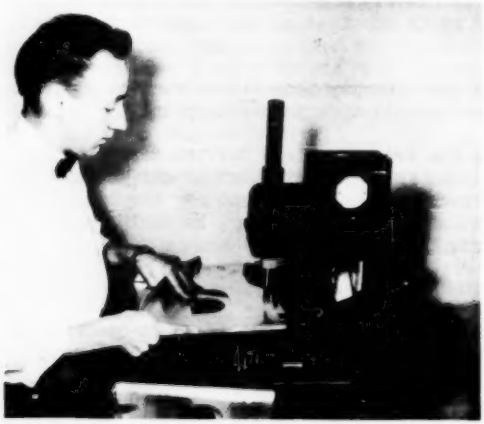


Fig. (5) Automatic Hardness Certifier

The heart of this machine is a Barcol hardness tester, the checking and stamping being tied in through a system of relays and solenoids. Being basically a comparator, it is, of course, imperative that the operator have available and use on a planned basis a representative group of hardness standards to insure absolute accuracy.

Sampling inspection techniques are used extensively throughout our operations, whenever the characteristic so accepted is proved by subsequent checks or operations. We have found the use of these sampling techniques to be particularly profitable in Receiving Inspection for the initial acceptance of such items as aircraft hardware, heat-treatment of castings and magnetic inspection of various items. The controls used in these sampling operations are contained in our Inspection Methods Bulletins and follow the requirements of Specification MIL Std. 105A with modifications.

In no case are sampling or spot checks used as the final acceptance of any characteristic affecting the structural or operational integrity of the airplane. For example, a particular type of actuator may be tested on a statistical sampling basis on the bench but each actuator must be 100% functionally tested after installation in the airplane. Unsatisfactory operation on the bench or the assembly line would be cause for immediate adoption of 100% bench check until the difficulty was found and corrected.

Approximately 250 Inspection Methods Bulletins are utilized to implement the Receiving Inspection control of general and specific

classes of material and equipment. Typical examples of such Inspection Methods Bulletins which employ use of statistical sampling tables are shown in Figures 6 and 7.

QUALITY CONTROL DEPT. FAIRCHILD AIRCRAFT HAGERSTOWN, MD.

INSPECTION METHODS BULLETIN

RECEIVING

NO. R-16 REV. III DATE 4/25/52 PAGE 1 OF 1

TITLE AIRCRAFT HARDWARE (NUTS, BOLTS, SCREWS, PINS AND CLAMPS), RECEIVING INSPECTION OF

Supersedes and cancels IMB-R-16, Rev. II, dated 12/11/51. Delete and destroy.

I. PURPOSE:

To outline receiving inspection requirements for subject components.

II. REFERENCES:

1. AN, NAS, AND, FSD Standards Drawings.
2. MIL-STD-105A (modified) - "Sampling Procedures and Tables for Inspection by Attributes"
3. DOI-74-23 - "Government Source Inspection of Mill Run Lots of Standard AN Hardware, Parts and Equipment".
4. IMB-G-28 - "Hardness Testing of AN STeel Bolts"
5. MPS-PR-21 - "Use of Magne Gage".
6. T.O. 29-1-8 - "Limitation on the Reuse and the Replating of Non-Metallic Insert Self-Locking Nuts"

III. INSPECTION: (Refer to Statistical Quality Control Chart)

1. Check for government source inspection and/or statements.
2. Check to assure that part conforms to applicable standards drawings, with respect to dimensions, and hardness.
3. Check for any imperfection in threads.
4. Check, visually, general condition of parts.
5. A single piece shall be selected from each container for finish check per Reference 5 where source inspection has not been accomplished.
6. On bolts 5/16", by 24 thread (and over) drilled for cotter pins; make certain that all protruding burrs, that may damage the fiber nut, have been removed.
7. Replated (Salvaged) Nuts containing non-metallic inserts are not acceptable.

NOTE: Refer to MPS-PR-18 for regulations governing magnetic inspection.

STATISTICAL QUALITY CONTROL CHART

This application of sampling inspection is applied to certain (critical) characteristics for stock control or in process control only. Sampling is conducted on these items because they have been inspected previously and will receive a final inspection at a later stage of assembly or functional test.

MIL-STD-105A (modified), Table IV, AQL 2.5

Lot Size	Sample	Ac	Re
2 - 7	All	0	1
8 - 65	7	0	1
66 - 180	25	1	2
181 - 300	35	2	3
301 - 500	50	3	4
501 - 800	75	4	5
801 - 1300	110	6	7
1301 - 3200	150	8	9
3201 - 8000	225	11	12
8001 - 22000	300	14	15

When number of defects found in sample is equal to or less than acceptance number, the lot from which sample is drawn shall be accepted. When number of defects reaches rejection number, the lot will be rejected.

jb

Revised by: /s/ D. H. College

APPROVED /s/W.A.Shuping QUALITY CONTROL DEPT

/s/R. O. Riggle A/F QUALITY CONTROL

Fig. (6) Inspection Methods Bulletin R-16

The Inspection Methods Bulletin illustrated in Fig. (6) covers the receiving acceptance of aircraft hardware on a sampling basis. Fig. (7) shows a specific Inspection Methods Bulletin establishing the functional test and inspection procedure for an oxygen flow indicator, the bench testing of which is done on a sampling basis.

QUALITY CONTROL DEPT.
FAIRCHILD AIRCRAFT
HAGERSTOWN, MD.

INSPECTION METHODS BULLETIN
SPECIFIC

NO. S-163
REV.
DATE 9/18/51
PAGE 1 OF 1

TITLE INDICATOR, OXYGEN FLOW AN6029-1, K-1

I. PURPOSE:
To establish a functional testing and inspection procedure for subject instrument.

II. REFERENCES:
1. AAF Spec. 40926 - "Test; Aircraft Oxygen Equipment Functional (Low Pressure)

III. RESPONSIBILITY:
Instrument Test Room shall test and inspect subject part in accordance with FAD Statistical Quality Control Procedure as given in the following table. This application of sampling inspection is applied to certain (critical) characteristics for stock control or in process control only. Sampling is conducted on these items because they have been inspected previously and will receive a final inspection at a later stage of assembly or functional test."

Lot Size	Sample	Ac	Re
0 - 7	All	0	1
8 - 65	7	0	1
66 - 180	25	1	2
181 - 300	35	2	3
301 - 500	50	3	4

NOTE: 1. AQL 2.5. Inspection Level II (Ref. MIL-STD-105A)(Modified).
2. If no defective units in excess of acceptance number are found in any sample; entire lot represented by that sample less known defectives, is acceptable.
3. If, in any sample, defects are found, equalling or exceeding the Rejection Number, entire lot shall be rejected.

IV. RECORDING:
Accepted indicators shall be ink-stamped with an FAD acceptance stamp, and serial number recorded.

V. DISPOSITION:
Tested units found acceptable shall each be ink-stamped by inspection personnel with an FAD acceptance stamp. Rejected valves shall be recorded and disposed of as noted in SPB-870-26.

VI. EQUIPMENT:
1. Oxygen Test Stand
2. Breathing Oxygen Cylinder (2000 psi capacity).
3. Laboratory Test Gage.

VII. PROCEDURE:
1. Indicator shall be visually inspected to assure freedom from cracked case, broken glass cover, defective dial markings, and other imperfections which would render the unit unsatisfactory for installation in aircraft.
2. Connect indicators to oxygen test stand manifold, and apply oxygen pressure of between 6 to 7 psi, as read on test gage connected to manifold.
3. By opening and closing line valve at one end of manifold, indicator shutters shall be made to open and close without sticking or binding.

Revised by: /s/ G. F. Schlichter

APPROVED: /s/W. A. Shuping — QUALITY CONTROL DEPT.
/s/R. O. Riggle — A/C QUALITY CONTROL

Fig. (7) Inspection Methods Bulletin S-163

We have not mentioned the control of variables at the point of origin by means of $\bar{X}$ & R charts in their many variations. The major application of this type control appears to be at the "floor level" by the operator or the inspector assigned to the process under way. It has been our experience that the efficient application of such controls is largely dependent on the rate of production of identical items. Most applications must involve production at least in the thousands and the decision to use such controls involves the balancing of the cost against the prospective benefits to be derived.

Due to the large variety of different parts being processed in one shop and the comparatively small number of pieces manufactured on any one order, the extensive use of $\bar{X}$ & R charts has been found neither economical nor practical. This technique, however, has been used occasionally on large quantities of small parts, such as bushings, to

familiarize our personnel with this tool so it may be applied should the need develop in the future.

P CHART (FRACTION DEFECTIVE)

Most routine inspection of manufactured airframe products is done by attributes, classifying each item inspected as either accepted or rejected (with possible further division of rejects into scrap and rework). This statement applies both to 100% inspection and sampling inspection. In such inspection, it has been common practice to make a record of the number of items rejected. This information has always been of interest to various levels of management and used more or less to guide corrective action. However, as the size of an operation increases and the volume of the information increases in the same proportion, it becomes impractical and often unnecessary to review the details of such reports at higher management levels. For this reason, we have designed a system of quality control reports and charts to cover the areas of responsibility of various levels of supervision. The Flow Chart, Fig. (8), illustrates how the information flows from daily or weekly reports at the first line supervisory level to the overall summary reports used for management control. Such a system, while it may appear cumbersome at first glance, actually summarizes for each supervisor the data which allows him to take most effective action without having to comb through voluminous reports with which he is not directly concerned. Such a system also places a control on the reports themselves to avoid duplication and excessive distribution. Throughout our system, we try to keep in mind that this quality control data is taken as a basis for doing something. If such data cannot be used by a supervisor as a basis for action, then we do not take up his time by requiring him to review it.

In analyzing the Parts Manufacturing stage of our operations, we make extensive use of the "p" or Fraction Defective chart in a modified form. The facts obtained for purposes of a "P" chart contribute information that should always be available to the management but frequently such information is not compiled in a manner which allows it to be reviewed and used to best advantage. The usual difficulty is that although numbers of defectives may have been regularly recorded, no record shows the numbers inspected for each group of defectives. To compare quality levels at different times, it is necessary to know the fraction defective at the various times. Every fraction has a numerator and a denominator. One great advantage of the "P" chart is that it requires the denominator as well as the numerator to be recorded and thus supplies information to management regarding current quality level which may fairly be compared with previous quality levels computed on the same basis. This statement holds true regardless of changes in production volume.

In our plant where there are a number of shop departments using many control charts for the fraction defective as mentioned, it has proven desirable to prepare charts which summarize all such charts under the area of supervision covered by each assistant superintendant. Such summary charts enable top executives to quickly spot the trouble areas without examining individual "P" charts.

In our assembly operations, we also make use of "P" charts to evaluate the defectives involving the use of Material Review actions. Here we use as the denominator the very handy performance figure of the

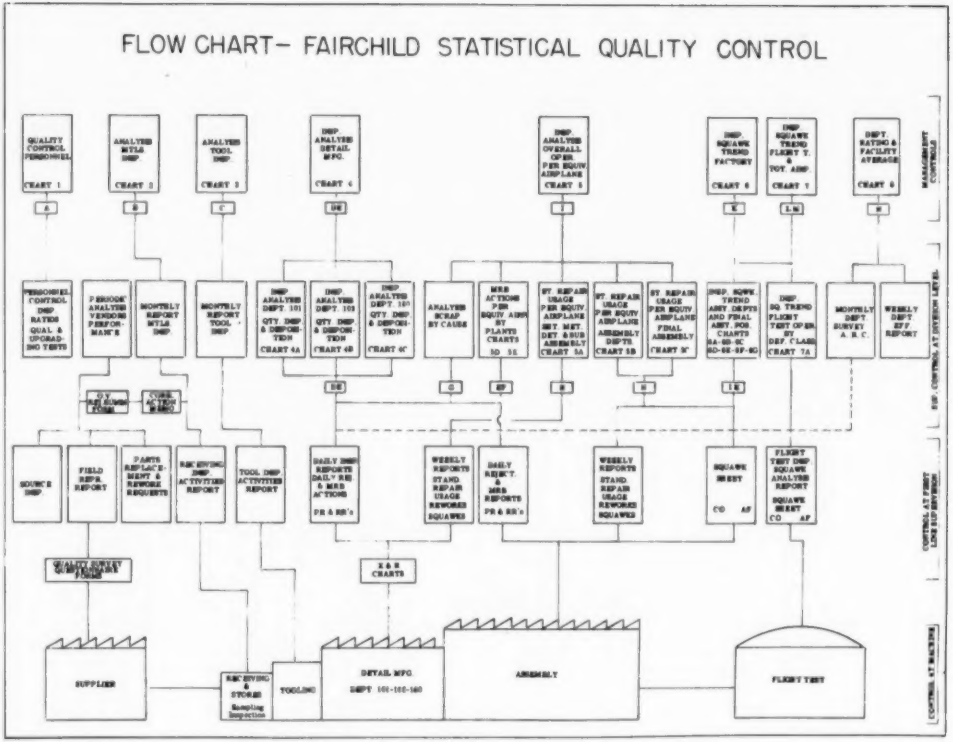

FLOW CHART– FAIRCHILD STATISTICAL QUALITY CONTROL
QUALITY CONTROL PERSONNEL
CHART 1
ANALYSIS MTLS INSP.
CHART 2
ANALYSIS TOOL INSP.
CHART 3
INSP. ANALYSIS DETAIL MFG.
CHART 4
INSP. ANALYSIS OVERALL OF REJ. PER EQUIV AIRPLANE
CHART 5
INSP. SQUAWK TREND FACTORY
CHART 6
INSP. SQUAWK TREND FLIGHT T. & TOT. ASSY
CHART 7
DEPT. RATING & FACILITY AVERAGE
CHART 8
PERSONNEL CONTROL INSP RATING QUAL & UPGRAD-ING TESTS
PERSONN ANALYSIS VENDORS PERFOR-MANCE
MONTHLY REPORT MTLS INSP.
MONTHLY REPORT TOOL INSP.
INSP. ANALYSIS DEPT. 101 QTY. INSP & DISPOSI-TION
CHART 4A
INSP. ANALYSIS DEPT. 103 QTY. INSP & DISPOSI-TION
CHART 4B
INSP. ANALYSIS DEPT. 105 QTY. INSP & DISPOSI-TION
CHART 4C
ANALYSIS SCRAP BY CAUSE
MRB ACTIONS PER EQUIV AIRP BY PLANTS CHARTS
5D 5E
ST. REPAIR USAGE PER EQUIV AIRPLANE INT. MET. DEPT. & BOIS ASSEMBLY
CHART 5A
ST. REPAIR USAGE PER EQUIV AIRPLANE ASSEMBLY DEPTS.
CHART 5B
ST. REPAIR USAGE PER EQUIV AIRPLANE FINAL ASSEMBLY
CHART 5C
INSP. SQUAW. TREND ASSY. DEPTS AND FINAL ASY. POR CHARTS
6A-6B-6C 6D-6E-6F-6G
INSP. SQ. TREND FLIGHT TEST OPER. DEPT. DEF. CLASS.
CHART 7A
MONTHLY DEPT. SURVEY A. B. C.
WEEKLY DEPT. EFF. REPORT
SOURCE INSP.
FIELD REPR. REPORT
PARTS REPLACE-MENT & REWORK REQUESTS
REC'DING INSP. ACTIVITIES REPORT
TOOL INSP. ACTIVITIES REPORT
DAILY INSP. REPORTS DAILY REJ & MRB ACTIONS
PR & RR's
WEEKLY REPORTS STAND. REPAIR USAGE REWORKED SQUAWKS
DAILY REJECT. & MRB REPORTS
PR & RR's
WEEKLY REPORTS STAND. REPAIR USAGE REWORKED SQUAWKS
SQUAWK SHEET
CO AP
FLIGHT TEST INSP SQUAWK ANALYSIS REPORT
SQUAWK SHEET
CO AP
X & R CHARTS
SUPPLIER
RECEIVING & STORES
TOOLING
DETAIL MFG
ASSEMBLY
FLIGHT TEST
MANAGEMENT CONTROLS
CONTROL AT DIVISION LEVEL
CONTROL AT FIRST LINE SUPERVISION
CONTROL AT SOURCE

Fig. (8) Flow Chart - Statistical Quality Control

"departmental equivalent airplane production".

Every operation done by a department is assigned a manhour standard. The total Standard manhours to make one airplane is, of course, known and the department output in any given period of time is evaluated in terms of equivalent airplanes. This is the figure we use as the common denominator when comparing quality performance between departments or for the same department at different times.

Fig. (9) shows a typical example of the "P" chart used by Management for overall control of Machine Shop operations. It will be noted that we plot the total quantities of parts inspected for each 15 day period on the upper portion of the chart and in the middle section is shown the dispositions of the material inspected in percentage, i.e. "Accepted on First Inspection, Return for Rework and Rejected". The lower section of the chart shows the percentage of parts scrapped to the quantity inspected. Similar charts are maintained for each individual detail manufacturing department. Other "P" type charts using percentage disposition are employed to portray the trend of operations in Receiving Inspection and Tool Inspection. The charts have proven valuable to supervision and to management as a guide to where corrective action is required.

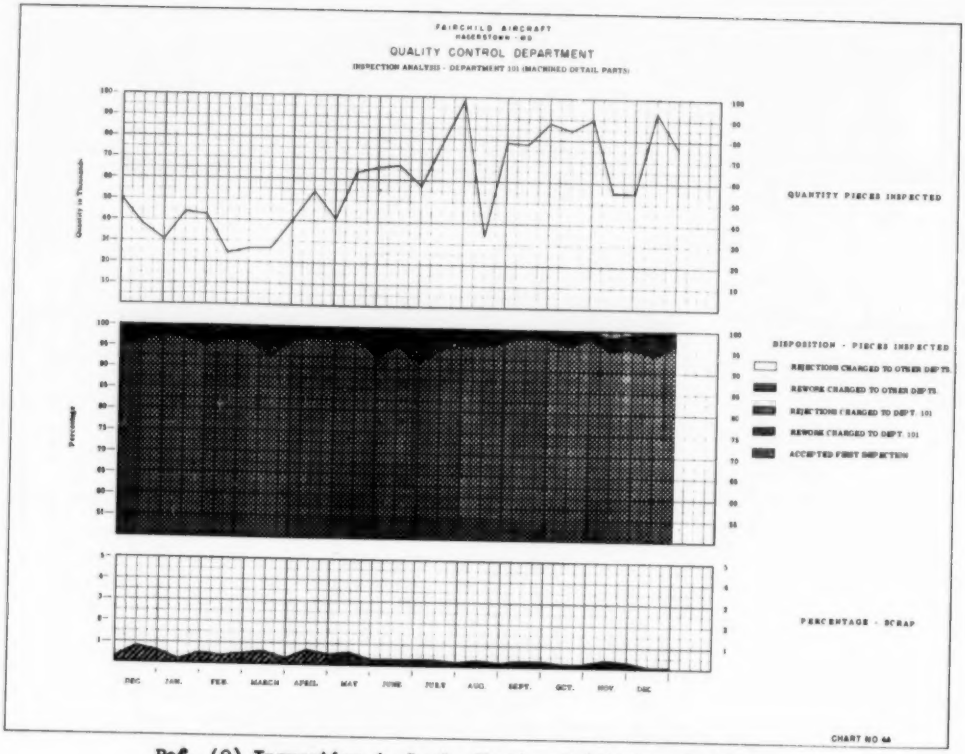
C CHART (DEFECTS PER UNIT)

In our assembly areas we make extensive use of modified "C" charts showing defects per unit. Such charts are posted in the assembly areas as a psychological aid to improvement in workmanship standards. Overall charts are available for supervision and management of factory and Flight Test operations. See Fig. (10)

If you will examine the "C" chart covering evaluation of our Flight Test operations, Fig. (11), you will note that it is based on the premise that some defects are more serious than others. This system provides for weighing of the defects found according to a scale that measures their relative seriousness. Again it should be pointed out that the value of these charts is absolutely dependent on rigidly defined standards of acceptance and a thorough knowledge of their application by inspection personnel.

GENERAL

At Fairchild we are firmly convinced that many of the things that bear on the quality of the product are not actually delivered to the customer. Housekeeping, parts handling and even performance against a budget and to a schedule are in a broad sense a part of Quality Control. With this in mind, we have a group composed of production, quality control and customer representatives constantly surveying our plant and analyzing reports to present what we believe is the ultimate statistical control to the general management each month. While a detailed discussion of their operations would be out of place here, if you will refer to Fig. (12), you will see a typical summary quality rating chart for the various departments and for the facility. This data represents the compilation of information from all sources including a survey of the physical operations. Such reports are used in evaluation of supervisory performance and to point out our trouble spots. Needless to say, such reports lend considerable interest to all of the components from which they are summarized.



Ref. (9) Inspection Analysis Chart - Detail Manufacturing

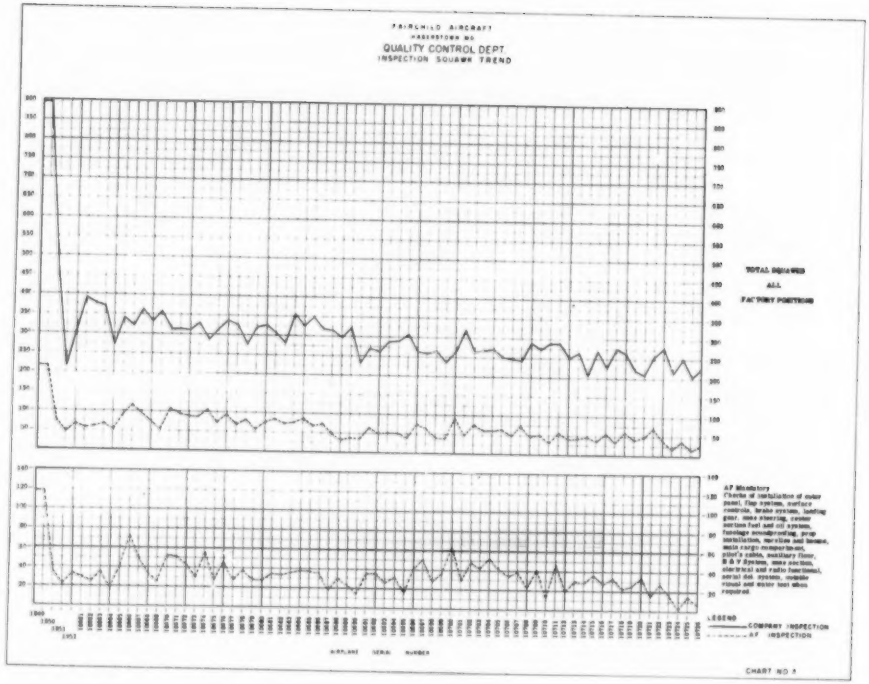


Fig. (10) Inspection Squawk Trend - Factory

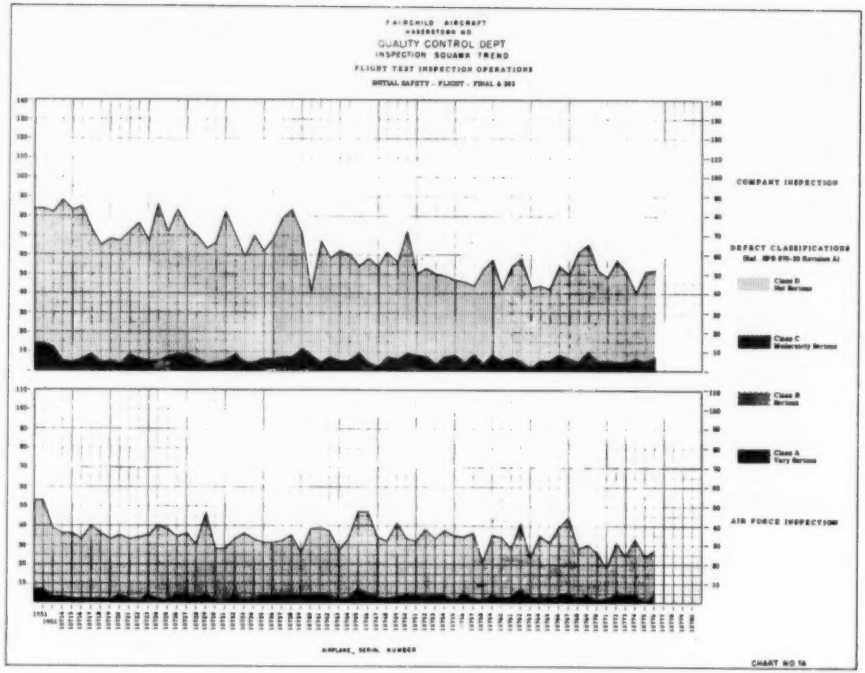


Fig. (11) Inspection Squawk Trend - Flight Test

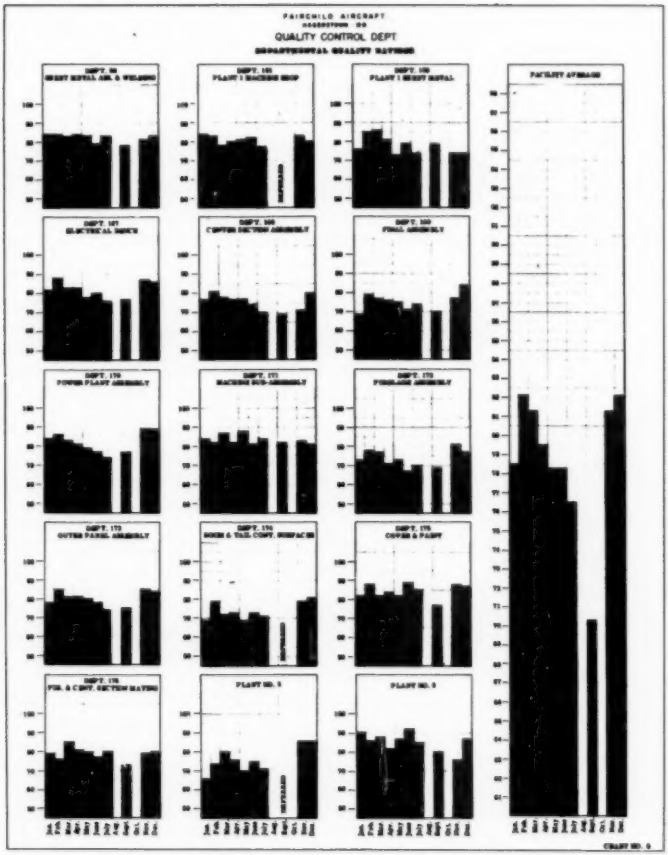


Fig. (12) Departmental Rating Chart

CONCLUSION

In conclusion, may we again state that Statistical Quality Control is a tool that:

Is only as good as data on which it is based.
Is only effective when it is used for action.
Is only efficient where it reduces cost or improves quality.

All of these things involve the personnel on the job. A good tool in the hands of the untrained and inexperienced is completely ineffective. A good tool used by skilled employees produces the best quality most efficiently.

REFERENCE:

Grant, E. L., "Statistical Quality Control", McGraw Hill Co. New York, N. Y. 1946

MILITARY INSPECTION PROCEDURES OF VEHICLES AND PARTS

J. P. Roche
Ordnance Corps
Detroit Arsenal

A large part of the Army purchasing mission is accomplished by the Ordnance Corps. The magnitude of this mission makes it highly imperative that procurement be accomplished swiftly, effectively, and to the best interests of the Government. Although the objectives of the procurement program must be expeditiously attained, economy and productive efficiency cannot be ignored. We are committed to obtain the maximum use of funds, and to secure the largest possible quantity of supplies of a requisite quality which can be purchased with those funds.

The practice of inspecting military supplies purchased from contractors is long established. The contract for the first naval vessels built for the Government after the Revolutionary War contained specific provisions for inspection of the vessels during construction, and for inspection of their armaments. Records show that the first batch of guns submitted by one manufacturer was rejected because the guns failed to meet specifications and firing tests provided in the contract.

From these early beginnings, inspection has continued to play an important role in the procurement of supplies and equipment. Each of the three major elements, namely purchasing, expediting, and inspection, has a definite and essential place in the supply program, and all must be coordinated into a smoothly working team for the effective procurement of military supplies.

Government inspection is required because, first, the consequences of defective material, incorrect material, etc., are sufficiently serious to make inspection a necessity, and second, Government inspection, rather than sole dependence upon the contractor's own inspection and/or warranty, is necessary to prevent acceptance of defective material.

The acceptance of defective supplies for military purposes is likely to have more serious consequences than the acceptance of defective items for normal civilian use. This is particularly true in time of war. Military supplies must be serviceable when they arrive at the point where they are to be used. There is seldom time to repair items found to be inoperative or unserviceable at the time the shipping container is opened, even if repair facilities are available. Worse still is the failure of equipment, without warning, after a short period of use. The individual serviceman's life and the success of large-scale military operations depend upon the adequacy of military equipment. When defective or substandard products are accepted in time of war, the consequences may be catastrophic. Deteriorated foods which caused the death of many soldiers during the Spanish-American War is a case in point. The peacetime storage of defective supplies over long periods for use in an emergency represents an equally serious situation. Equipment which is not in serviceable condition, but which is thought to be, is worse than no equipment at all, since it promotes a false sense of security that may be disastrous.

Defective manufacture of the desired items is not the only condition that must be avoided. Improper preservation, improper packing

and marking of shipments, and incomplete identification of otherwise satisfactory equipment can also have serious consequences.

The exceptionally rough handling which military equipment receives as it is transported to the theater of operations, and the unusually severe conditions under which it is stored before it is used, makes the factors of preservation and packing assume major importance in the inspection of military equipment. In some instances, items must be individually packaged so that the external container may be opened and part of the contents removed without exposing the remainder to spoilage. Testing, packaging, and shipping an item serve no purpose if the item arrives in an inoperable condition or if it becomes damaged in storage.

The Industrial Division, Office of the Chief of Ordnance, has been assigned the staff function of "formulation and supervision of policies and procedures relating to inspection of materiel procured by the Ordnance Corps".

The Production Service Branch, in connection with inspection of materiel procured by the Ordnance Corps, formulates and supervises basic policies designed to eliminate unnecessary inspection duplication and to produce uniformity and simplification of practices. It provides general principles and patterns for application of sampling and Quality Control; and is responsible for the staff supervision of writing, arranging for publication, and maintaining the Ordnance Inspection Manual.

The Ammunition, Artillery and Small Arms Centers have been assigned the operational functions of formulating and supervising inspection policies and procedures relating to their commodity requirements.

The National Tank Automotive Inspection Office, Ordnance Tank Automotive Center, located at the Detroit Arsenal, has been assigned the operational function of formulating and supervising inspection policies and procedures relating to all tanks, and automotive vehicles and equipment procured by the Ordnance Corps. This office is responsible for quantity and quality of tank-automotive items procured under the current Army Supply Program, and determines mandatory inspection requirements including quality control procedures, standards of acceptability, and defect classifications for all items procured by the Ordnance Tank Automotive Center. These policies and procedures must be complied with by proving grounds, arsenals, and Ordnance Districts.

Each Ordnance District has been assigned the operational function of planning and determining the number of inspectors necessary to successfully accomplish the inspection of all items procured within the District, and of accomplishing inspection in accordance with these plans. Planning and carrying out the inspection job is a fundamental requirement on all contracts.

World conditions and the position of this country in world affairs makes the building and maintenance of a formidable military machine mandatory. The utmost utilization of all resources is imperative. Among the most critical elements in the expansion of the military potential of our country are the procurement, training and effective utilization of its manpower resources. From the inception of raw materials to their end use at the fighting front the Department of the Army emphasizes the necessity for the progressive, efficient and intelligent employment of

personnel.

During World Warr II, more than 50 percent of the total number of employees engaged in carrying through the government procurement requirements were assigned to inspection organizations. During such periods of intensified buying, the availability of highly trained inspection personnel is negligible. Plans must therefore be devised to the end that the minimum amount of inspection necessary to give full protection to the interests of the government be provided, and the most effective utilization of the available inspection manpower be accomplished. These plans must be so devised that responsibility can be delegated to less experienced inspectors by prescribing specific inspection methods and procedures. Plans incorporating optimum use of Statistical Quality Control make possible the acceptance of quality material while maintaining the utmost economy with minimum personnel.

The Ordnance Corps of the Army was the first branch of the armed services to use statistical sampling plans for acceptance of material. In 1942 sampling procedures and tables for inspection by attributes were compiled under the direction of engineers from Bell Telephone Laboratories for the Army Service Forces. In 1945 tables and procedures were developed for the Navy by the Statistical Research Group of Columbia University. After unification of the armed services the Navy's tables were adopted by the Department of Defense as a joint army and navy specification, JAN-STD-105. The present military standard, MIL-STD-105A superseded JAN-STD-105 in September 1950 and at present is again being revised to include tables for sampling by variables.

Government contracts and specifications require the submittal of only acceptable material for acceptance inspection. Wheeled vehicles must be 100 percent inspected and road tested by the vendor before submittal. General specifications for automotive vehicles and replacement parts require a contractor to inspect for and repair a defect on all vehicles in his possession, whether completely or partially assembled, when that defect is of such a nature that, in the opinion of the Ordnance Inspector will require rejection of that vehicle by the Government. The contractual right of lot acceptance or rejection was based upon this requirement.

Willys Overland Motors Inc., producers for the government of M-38 A1 Jeeps, were contacted relative to initiating sampling acceptance of their product. This being a new experiment in vehicle acceptance, the co-operation of all concerned parties was needed.

Government acceptance had previously been accomplished by 100 percent re-inspection and re-road-test by Ordnance Personnel. Despite contractor precautions, this method, because of the human element involved, has a normal tendency to allow passing of defective material to the Ordnance inspectors. The basic purpose of government inspection should only be the critical examination of the contractor's inspection, road test and repair process. Acceptance through sampling by attributes affords assurance of requisite quality, while placing the burden of finding and repairing defective items on the contractor's personnel.

Data for currently purchased vehicles was available in the form of inspection check sheets for many months of previous production. From this data, the recurrence of different classes of defects were tabu-

lated for the purpose of establishing mutually agreeable quality levels.

A quality control inspection procedure was developed which contains a classification of defects with acceptable quality levels for each classification. The extent of non-conformance of product is expressed in terms of number of defects per hundred vehicles. The single sampling plan was chosen. The lot size was established at sixty, due to the space shortage at the contractor's plant. The inspector is instructed to select at random ten sample vehicles from each lot. Each sample vehicle is to be inspected completely before, during, and after a two mile road test in accordance with the check sheet previously used for 100 per cent inspection. The inspector is cautioned to maintain sample and lot identity and to completely inspect the lot sample. Provisions for repairs and inspection of re-submitted lots is outlined. The type of charts to be maintained and the distribution of such finalized charts is stated.

Defects are classified as closely as possible in accordance with definitions included in MIL-STD-105A. It was found necessary to add two classifications not found in this standard; Special defects, defined as: "Those defects which, while not critical in themselves, indicate a serious breakdown in production procedure and control", and incidental defects, defined as: "Those defects which are deviations from good workmanship or from specified requirements of the type that would probably have no effect on the functioning, assembly, use, interchangeability, repair, or life of the vehicle".

The classification of defects with acceptable quality levels was established as follows:

ACCEPTABLE QUALITY LEVELS AND CLASSIFICATION
OF DEFECTS

Truck, 1/4 Ton 4x4, M38A1

SPECIAL AQL - 0 Def/100 Vehicles: The defects listed in this classification will be inspected 100 percent by the Ordnance Inspector. This inspection will be made on every vehicle and will be the final operation on each lot of vehicles before final acceptance or rejection is made.

SPECIAL DEFECTS - 0 Def/100 Units

Defect No.

A. Fuel leaks - Drips
B. Coolant leaks - Drips
C. Lubrication leaks - Drips
D. Brake Fluid leaks - Any
E. Coolant missing
F. Sheet metal damage
G. Foot brake inoperative
H. Steering mechanism loose or binding
I. Components omitted or missing
J. Any apparent serious malfunction or damaged part

MAJOR DEFECTS - 15 Def/100 Vehicles

Defect No.

1. Steering Knuckles and linkage - Loose or excessive play
2. No shift or unusual noise in power train
3. Excessive Vibration
4. Clutch - slippage, drag, chatter or maladjustment
5. Oil Pressure - low or gage inoperative
6. Engine Heat - high or gage inoperative
7. Brake, foot - Grab, chatter, unequal pull or maladjustment
8. Excessive heating - All units in power train
9. Lights - Complete circuit inoperative
10. Glass, windshield and windows - fogged, cracked or broken
11. Wipers, windshield - inoperative
12. Vacuum test, engine
13. Fuel tank - not strapped securely
14. Pintle hook - Inoperative or Improperly mounted
15. Hoists, winches and Etc. Where specified - Improper mounting, assembly or adjustment; Damage

MINOR DEFECTS - 150 Def/100 Vehicles

Defect No.

51. Pedal or linkage adjustment required - Hand & foot brake; clutch; Accelerator
52. Fording Controls - Inoperative
53. Instruments inaccurate - Ammeter, fuel, speedometer, etc.
54. Unusual Engine noise
55. Engine tune up required
56. Improper lubrication level, Engine, Transmission, etc.
57. Improper coolant level
58. Horn - inoperative
59. Pulley, Crankshaft - wobble or misaligned
60. Shock Absorbers - Improper mounting or leaks
61. Batteries, Cables, acid level - defective or improper otherwise
62. Wiring and tubing - defective or improperly mounted
63. Fan assembly, blades, belts - Improper alignment or otherwise defective
64. Generator - defective mounting or alignment
65. Distributor - Improper wiring (includes plugs and filter)
66. Lifting loop (Engine) - damaged
67. Springs - improper assembly or deflection
68. Fuel pump - improper mounting or assembly; Damaged
69. Carburetor - improper mounting or assembly; Damaged
70. Power Train - improper mounting or assembly; Damaged
71. Radiator - improper mounting or assembly; Damaged
72. Oil Filter - improper mounting or assembly; Damaged
73. Master Cylinder - improper mounting or assembly; Damaged
74. Water Pump - improper mounting or assembly; Damaged, inoperative; Leak
75. Starter - improper mounting or assembly; Damaged
76. Iginition System - improper mounting or assembly; Damaged
77. Wheels - improper mounting or assembly; Damaged

INCIDENTAL DEFECTS - 400 Def/100 Vehicles

Defect No.

101. Body and Hood and items mounted thereon - Damage, Improper fits, Assembly welds, seals and adjustment
102. Sheet metal defects
103. Paint defects
104. Trim defects
105. Lifting shackles - Improper mounting
106. Tires - damage or improper assembly
107. Controls & Switches - Improper mounting, clearance, or condition
108. Decals & Instruction plates - Improper position or condition
109. Manifolds, Air Cleaner, crankcase breather - Improper assembly or damaged
110. Vent lines for fording - Improper assembly or damaged
111. Frame - Improper Assembly - Loose bolts or rivets - Broken welds
112. Bumpers - Improper mounting or assembly; Damaged
113. Muffler and Pipes - Improper mounting or assembly; Damaged
114. Light bulb - Burned out or cross wired
115. Grease fittings - Missing, damaged or improperly installed

After completion of road test and final inspection by the Ordnance Inspector, each vehicle is presented to the Ordnance acceptance inspector where 100 per cent inspection is made of brakes, steering mechanism, general functioning of vehicle and leaks shown as special defects. The other special defects, such as damage, missing parts or malfunction, are placed in this category to allow repair or replacement of obvious deficiencies even though acceptance of the lot has been made through sampling. Since all contractors desire production of a quality item, there have been no objections to these inclusions.

During the first period of inspection under this procedure, eighty eight lots were inspected. The process average for major defects was 8.86 defects per hundred units, for minor defects 127.16 defects per hundred units, for incidental defects 335 defects per hundred units. Two lots had been rejected for minor defects and four lots for incidental defects. These six rejected lots were less than was anticipated, especially during a "start-up" period. It will be noted that the number of defects per 100 vehicles is less than the acceptance quality level established by the Standard Inspection Plan. The savings in Ordnance inspection personnel amounted to a reduction of nineteen road testers and line inspectors previously maintained for 100 percent inspection. Data show a drop from an average of 8.3 defects on each vehicle submitted, prior to initiation of a sampling plan, to an average of 4.66 defects per unit under the present system of inspection. Figures at the present time show 3.4 major defects per hundred units, 73.2 minor defects per hundred units and 389.2 incidental defects per hundred units. This shows decided improvement in regards to major and minor defects. Although not apparent, the improvement in incidental defects is definitely there. As the more apparent defects disappear, the inspectors become more critical. The scope of the incidental category permits varied degrees of criticism.

The government inspectors were very frank to admit their skepticism,

which existed during the early stages of this experiment, but they are equally frank about acknowledging the beneficial results of the program as organized. Recognition should be made of the excellent co-operation given during the missionary period by Ordnance road testers and line inspectors.

In the analysis of data so far received we find a low, elongated distribution curve. Within this curve we find four distributions, one for each inspector, showing that inspectors apply varied degrees of judgement. We hope, by the progressive development of pictorial diagrams, to standardize inspection's thinking relative to item deficiencies.

With each purchase of vehicles made by the government, an order is placed for spare parts necessary to maintain those vehicles for an estimated period of ten years. These parts are purchased concurrently in sets with each predetermined group of vehicles. They are purchased as separate items, and packaged and stored for future use.

The statistical sampling of finished vehicles gives only a superficial assurance of quality. The statistical sampling of concurrently manufactured spare parts for those vehicles, with proper co-ordination between spare part and production inspectors, gives an adequate quality assurance for the complete, complex unit. There are thousands of parts in each vehicle and thousands of component inspection procedures must be written. We expect to accomplish this task by having each Ordnance District prepare and use their own plans, pending submission of such plans to the National Tank Automotive Inspection Office for review, correction, approval and eventual standardization.

The importance of cleaning, preserving and packing of replacement parts is evident. The cost of this process often exceeds the cost of manufacture, and many times has been the cause of adverse criticism to the government. The lives of our servicemen and preservation of our way of life can depend on the receipt of unimpaired and undamaged material at the required place. In accomplishing this, any intelligent expenditure of funds is warranted. There are over six hundred (600) specifications governing the process of packaging and packing. It is highly imperative that standard inspection procedures, incorporating statistical quality control plans, be prepared for the purpose of minimizing the inspection work-load, assuring requisite quality, attaining maximum economy and standardizing specification interpretation. The preparation of these procedures has been initiated.

The three (3) basic phases of vehicle inspection become:

1. Completed vehicle inspection
2. Spare part inspection
3. Packaging and packing inspection

Relative to the first phase, the experiment of sampling acceptance, put into effect at the Willys Overland Motor Company by the Cleveland Ordnance District, has developed into a general quality control inspection procedure for automotive vehicles. Using this procedure as a format, the Detroit Ordnance District has established and put into effect sampling procedures for trucks being procured in the Detroit Area. A tentative general quality control inspection procedure for medium tanks

has been established using the Willys Overland plan as a pattern and is now in use in the experimental stage. A general procedure for trailers will be established in the near future. At the present time it must suffice to say that sampling inspection plans have been initiated.

Reports of sampling results, relative to all phases of vehicle inspection, will be submitted to the National Tank Automotive Inspection Office for compilation of master control charts. Analytical reports will be prepared from these charts and forwarded to the Chief, Key Inspector Division for presentation to the Officer-in-Charge of Ordnance Department Industry Integrating Committees, thereby standardizing manufacturing processes to the end that troublesome conditions are eliminated.

We in the National Tank Automotive Inspection operation realize that many problems must be solved as we confidently anticipate a successful "New Era" in the acceptance for the Ordnance Corps of complete wheeled and tracked vehicles by means of Statistical Quality Control methods.

Some portions of this paper have been taken from the following government publications:

"Introduction to Procurement Inspection"
"M608 - Inspection Manual"
"Circular for Manpower Utilization"

The following persons were involved in the preparation of this paper and the Quality Control Inspection Procedure for Willys Overland Jeeps:

Lt. Col A. N. Bray, O.I.C., National Tank Automotive Inspection Office
H. Sprengel, Chief, National Tank Automotive Inspection Office
R. W. Clark, Chief, Key Inspector Division, National Tank Automotive Inspection Office
L. Canfield, Chief, Inspection Division, Cleveland Ordnance District
W. P. Jones, Chief, Quality Control Branch, Cleveland Ordnance District
E. Martz, Resident Inspector of Ordnance, Willys Overland Motor Company
M. Hutchinson, Asst. Resident Inspector of Ordnance, Willys Overland Motor Company
N. Meagley, Willys Overland Motor Company

The necessity for the establishment of lot sizes in accordance with availability of storage space, and the tendency to pass questionable material to Ordnance inspectors, as previously stated in this paper, are conditions common to all tank and automotive producers throughout industry, and are not to be construed as criticism of the Willys Overland Motor Company. The efforts of this company under the personal direction of Mr. R. R. Rausch, their Vice President and Executive Assistant to the President, in the pioneering and application of this program, are hereby acknowledged.

CHI SQUARE TESTS OF INDEPENDENCE - EXPOSITION AND COMMENT

Acheson J. Duncan
The Johns Hopkins University

1. Description of a χ^2 Test

In industrial research tests of independence are of prime importance. When the items under investigation are classified according to their attributes and the frequency entered in a contingency table such as illustrated by Figure 1, it is customary to test for independence by running a χ^2 test. It is the purpose of this article to discuss these χ^2 tests of independence.

Figure 1

A Form for the Study of Independence

		Classification I Category				
		A	B	C	D	Total
Classification II						
Category	a	—	—	—	—	—
	b	—	—	—	—	—
	c	—	—	—	—	—
Total		—	—	—	—	—

To keep the discussion concrete consider the data of Table 1. These pertain to failures of vacuum tubes used during World War II in V T fuses. The data were derived in the course of attempts to standardize test conditions for the fuses and provided an answer to the question: Is the distribution of tubes by type of failure the same when the tube is located in the top block of the shell as when it is placed in the bottom block? In other words, is type of failure (Classification I) independent of position in the shell (Classification II)?

In general a test of independence consists of two steps: 1. determining how the data would be expected to be distributed if the two classifications were independent and 2. comparing the actual distribution with the expected. Let us illustrate the first step with respect to the data of Table 1. If the distribution of type of failure is independent of the position of the tube in the shell, we would expect the proportions of the tubes falling in categories A, B, and C to be the same for the top block as for the bottom block of the shell. In other words, we would expect the frequencies in the row marked "Top Block" to be proportional to the frequencies in the row marked "Bottom Block", the factor of proportionality being the ratio of total failures in the bottom block to total failures in the top block. Another way of saying the same thing is that we would expect the 115 type A failures to be distributed in the same proportion as the total failures, viz., in the

Table 1

Failures of V. T. Vacuum Tubes*

Position in Shell	Type of Failure			Total Failures
	A	B	C	
Top Block	75	10	15	100
Bottom Block	40	30	10	80
Total	115	40	25	180

* Reference (3), p. 18.

ratio of 8 tubes in the bottom block for each 10 tubes in the top block. Likewise, we would expect the 40 type B failures to be distributed between bottom and top blocks in the ratio of 8 to 10, and also the 25 type C failures. The distributions expected on the hypothesis of independence would, therefore, be those given in Table 2.

Table 2

Frequencies of Failures for the Data of Table 1 That Would Have Been Expected* on the Hypothesis of Independence, Given the Marginal Totals

Position in Shell	Type of Failure			Total Failures
	A	B	C	
Top Block	63.9	22.2	13.9	100.0
Bottom Block	51.1	17.8	11.1	80.0
Total	115.0	40.0	25.0	180.0

* Expected in the sense of being an average over many samples. This is why the frequencies are not rounded off to whole numbers.

The data of Table 1 do not conform exactly to those of Table 2. The data of Table 1 are sample data, however, and allowance must be made for sampling fluctuations. What we would like to know is whether the results of Table 1 differ sufficiently from those of Table 2 to make the hypothesis of independence untenable? To put it another way, is such a sample result very unlikely if we assume the existence of independence and allow for sampling fluctuations? It is to answer such questions that the χ^2 test is employed.

A χ^2 test consists of comparing actual and expected frequencies.

The formula used is

$$(1) \qquad \chi^2 = \sum \frac{(F - f)^2}{f}$$

where F is the actual frequency of any cell, f is the frequency expected on the hypothesis of independence, and the sum is taken over all cells. For Tables 1 and 2, this formula yields the following:

$$\chi^2 = \sum \frac{(F - f)^2}{f} = \frac{(75 - 63.9)^2}{63.9} + \frac{(10 - 22.2)^2}{22.2} + \frac{(15 - 13.9)^2}{13.9}$$

$$+ \frac{(40 - 51.1)^2}{51.1} + \frac{(30 - 17.8)^2}{17.8} + \frac{(10 - 11.1)^2}{11.1}$$

$$= 1.93 + 6.70 + 0.09 + 2.41 + 8.36 + 0.11$$

$$= 19.60$$

The problem is how to evaluate this result.

According to statistical theory, if we should run many tests of the kind exemplified by Table 1 (not necessarily keeping the total tubes tested equal to 180, but always keeping the same classifications and, of course, making a random selection of tubes) and if the hypothesis of independence is actually true, then the sample values obtained by applying formula 1 to each set of sample data would tend to form a special frequency distribution called the χ^2 distribution. (Hence the name Chi-Square Test.) There are various χ^2 distributions, distinction among them being made on the basis of the so-called "degrees of freedom". This term will be explained later. For the moment let it be simply stated that for the problem in hand the degrees of freedom are 2. With this in mind note from Table 3 that, for a χ^2 distribution with 2 degrees of freedom, not more than 5% of the sample values of $\chi^2 = \sum \frac{(F - f)^2}{f}$ would, if the hypothesis of independence were true, exceed the value 5.99 nor more than 1% exceed the value 9.21.

Suppose now that in making a test of independence we are willing to run the risk of 0.05 of rejecting the hypothesis when it is true. In such a case, we will take the 0.05 point of the pertinent χ^2 distribution as a critical value for comparison with our sample result. If the sample value of χ^2 exceeds this critical value, we will reject the hypothesis; if not, we will accept it. In the problem in hand, the 0.05 point of the χ^2 distribution with 2 degrees of freedom is 5.99. The sample χ^2 is 19.60. Hence, we will reject the hypothesis of independence and conclude that type of failure is related to the position of the tube in the shell.

The above explains the technique of carrying out a χ^2 test of independence. Consider now the meaning and determination of the quantity called "degrees of freedom". For the data of Table 1, it was said that there were 2 degrees of freedom. By this it was meant that we had only 2 independent points of comparison between actual and expected frequencies. Why only 2, when formula 1 calls for 6 comparisons? The answer is that in computing the expected frequencies we made use of the

marginal totals, 100, 80, 115, 40, and 25, (as indicated above) in such a way that the actual and expected frequencies in every row and every column added up to the same marginal totals. Consequently we could only determine expected frequencies independently for two cells. All the other expected frequencies had to be such that the marginal totals were the same as for the actual frequencies. Hence we had only 2 "degrees of freedom". In general, if classification I has c categories and classification II has r categories, the degrees of freedom for the χ^2 test will be $(c-1)(r-1)$.

Table 3

The 0.05, 0.01, and 0.001 Points of the χ^2 Distribution*

Degrees of Freedom	Probability of a X^2 Equal to or Greater Than Listed Value is:		
	0.05	0.01	0.001
1	3.84	6.63	10.8
2	5.99	9.21	13.8
3	7.82	11.3	16.3
4	9.49	13.3	18.5
5	11.1	15.1	20.5
6	12.6	16.8	22.5
7	14.1	18.5	24.3
8	15.5	20.1	26.1
9	16.9	21.7	27.9
10	18.3	23.2	29.6
11	19.7	24.7	31.3
12	21.0	26.2	32.9
13	22.4	27.7	34.5
14	23.7	29.1	36.1
15	25.0	30.6	37.7

* The writer is indebted to R. A. Fisher and Frank Yates, and to Oliver and Boyd, Ltd., Edinburgh, for permission to reproduce the above from Table III of their book, Statistical Tables for Use in Biological, Agricultural and Medical Research.

2. Comments

a. The Level of Significance. The test of Section 1 was run, as we say, at a 0.05 level, the critical χ^2 being the 0.05 value. It is possible, and quite proper, of course, to run tests at other levels, say to take $\chi^2_{0.10}$ or $\chi^2_{0.01}$ or $\chi^2_{0.001}$ as critical values. In the first case, we would increase the risk of falsely rejecting the hypothesis of independence and in the other two we would reduce this risk. By way of comparison the selection of $\chi^2_{0.001}$ as a critical value is roughly the same as choosing 3σ limits on an $\bar{X}$ chart. Whether we choose the 0.10, 0.05, 0.01, or 0.001 point will depend on the problem in hand. If it is important to find even minor deviations from independence and the action based on such a finding does not lead to serious loss if we mistakenly

conclude there is dependence when the classifications are actually independent, then we would prefer the 0.05 and even possibly the 0.10 critical value. On the other hand, if it is not very important to discover minor deviations from independence, but very costly to conclude mistakenly that there is dependence, then the 0.001 critical point would be preferred. This is a question of risks and costs.

b. Adjustments because of the Approximate Nature of the χ^2 Test. The χ^2 test of independence is, in a probability sense, an approximate test. Just as the normal distribution or Poisson distribution is used as a labor saving device to compute approximate values for binomial probabilities, so the χ^2 distribution is used as a labor saving device to compute approximations to the exact probabilities theoretically called for by the problem. For this reason, it has been common practice in applying a χ^2 test of independence not to compare the actual and expected frequency of any cell unless the expected frequency for the cell is 10 (some say 5) or more. More recent research, (1)(2), suggests that when there is more than one degree of freedom, the χ^2 approximation is accurate enough for practical purposes provided that the smallest expected frequency is at least 2.

If the rule of 2 can not be met in a particular case, it may be circumvented by amalgamation of categories. With reference to Table 1, for example, if the expected number of Type C failures in the bottom block had turned out to be 1 instead of 11.1, then violation of the rule of 2 could have been avoided by amalgamating Type B and Type C failures into a single category that might have been called "Other Failures". Another possible solution would be to increase the total number of tests until the required expected frequency was obtained. The disadvantage of the first procedure is that the amalgamation of categories reduces the degrees of freedom of the test and hence lowers its sensitivity to actual deviations from independence. The disadvantage of the second procedure is the obvious one of increased expense.

If the data are so meager as to produce a number of small expected frequencies (possibly all may be small) and if it is considered to be too expensive to increase the total number of cases, it may still be possible to use the χ^2 test by correcting X^2 for continuity. (See reference (2).) If this does not sufficiently improve the approximation, the χ^2 test may have to be abandoned and resort made to the exact computation of probabilities (2). The difficulty with the latter is, of course, that it may require excessive labor.

c. The Effect of the Size of the Sample. Since the rule of 2 may be met by increasing the size of the sample, a word should be said about the relationship between the size of the sample and X^2. If a sample is doubled in size, say, and if the frequencies in the various cells of the contingency table maintain their relative proportions, then X^2 will also be doubled. (Cf. Formula 1.) For X^2 to remain unchanged, the differences between the new actual and expected frequencies will have to be reduced to the point that they are equal to $\sqrt{2}$ times the old differences instead of twice the old differences. This means the larger the sample the smaller the relative differences between actual and expected frequencies must be in order to have the same probability of occurrence. There is nothing strange about this. If we tossed 5 coins and got 3 heads, we would not be surprised in the least; in fact, this would be one of the more likely results. If we tossed 5000 coins, however, and

got 3000 heads, we would be inclined to think the coins were biased or had not been tossed in a random manner.

2d. The OC Curve for a χ^2 Test of Independence. The OC Curve for a χ^2 test of independence gives the probability of accepting the hypothesis of independence when in fact the classifications are not independent. Let us take as a measure of departure from independence the quantity

$$\lambda = \sum_{i=1}^{i=k} \frac{(m_i' - m_i)^2}{m_i} \tag{2}$$

where the m_i are the frequencies expected for the various cells on the hypothesis of independence and m_i' are frequencies expected on some other hypothesis. For such a definition of departure from independence, Evelyn Fix (6) has tabulated abscissa and ordinates for OC Curves for degrees of freedom from 1 to 100. A selected set of her results are presented in Table 4.

Table 4

Values of λ for which the Probability of Falsely Accepting the Hypothesis of Independence is 0.10* (Probability of Falsely Rejecting Hypothesis of Independence = α)

Degrees of Freedom	λ	
	$\alpha = 0.05$	$\alpha = 0.01$
1	10.509	14.879
2	12.655	17.427
3	14.172	19.248
4	15.405	20.737
5	16.470	22.033
6	17.419	23.187
7	18.284	24.238
8	19.083	25.211
9	19.829	26.122
10	20.532	26.981
20	26.132	33.852
30	30.379	39.074
40	33.941	43.461
50	37.069	47.312
70	42.48	53.99
100	49.29	62.39

* Abridged from Reference (6). Smaller λ 's yield higher probabilities of accepting the hypothesis of independence.

To illustrate the use of Table 4, consider a 2 x 2 contingency table in which 100 items are cross-classified. For example, suppose 100 VT vacuum tubes are cross-classified with respect to type of failure (A and Other Types) and position in shell, and suppose that the marginal totals are as follows:

Position in Shell	Type of Failure: A	Other	Total
Top			30
Bottom			70
	60	40	100

Then on the assumption of independence the expected frequencies would be

	A	Other	Total
Top	18	12	30
Bottom	42	28	70
	60	40	100

These expected frequencies are the m_i of formula 2.

Now from Table 4 it follows that we could actually have a dependence between position in shell and type of failure such that the actual expected values (the m_i' of formula 2) would differ from the frequencies we expected on the hypothesis of independence (the m_i) by as much as 7, i.e., we could have expected frequencies as follows:

	A	Other	Total
Top	11	19	30
Bottom	49	21	70
	60	40	100

and still have more than a 0.10 chance of accepting the hypothesis of independence.* If we should not like this result, we can increase the number of tubes to be tested. For this will increase λ (Cf. Formula 2) and reduce the probability of failing to detect the departure from independence.

In a practical problem it may be difficult to give an industrial significance to a particular value of λ. Consequently, the OC curve of a χ^2 test may not have as much meaning as does the OC curve for a sampling inspection plan. Nevertheless, consideration of the OC curve of the χ^2 test leads to consideration of the adequacy of the sample size for the purposes of the test and at least warns the user that if he accepts the hypothesis of independence when the sample size is small, he runs a sizeable risk of being wrong.

* For these data $\lambda = \sum \frac{(m_i' - m_i)^2}{m_i} = \frac{(11-18)^2}{18} + \frac{(19-12)^2}{12} + \frac{(49-42)^2}{42} + \frac{(21-28)^2}{28} = (7)^2 \left(\frac{1}{18} + \frac{1}{12} + \frac{1}{42} + \frac{1}{28}\right) = 9.7$ which is less than the λ in Table 4 for a probability of 0.10 and $\alpha = 0.05$ (d.f. = 1). A difference between the m_i' and m_i of 8 or more would have had less than a 0.10 chance of being detected.

e. The χ^2 Test as a Test of Difference Between Two Proportions. When the two classifications of a contingency table have only two categories each, the χ^2 test of independence is the mathematical equivalent of a test of a difference between two proportions. For example, with reference to Table 1, if there were only two types of failures, A and B, then the χ^2 testof independence becomes a test of whether the proportion of A failures in the top block is the same as the proportion of A failures in the bottom block.

When only two categories are involved and the frequencies are not large (say smallest expected frequency is less than 500), a more precise result is obtained (2) by taking

$$(3) \qquad X^2 = \sum \frac{(|F - f| - 0.5)^2}{f}$$

This subtraction of 0.5 is called the "continuity correction" for X^2 and is valid only for the four fold contingency table.

f. Combining of Independent X^2's. A χ^2 test of independence has the fortunate property that the results of several independent tests may be readily combined (7). For example, if another set of data pertaining to failures of VT vacuum tubes is analyzed in a manner similar to the analysis of the data of Table 1, then the two sample X^2 can be added and tested with degrees of freedom equal to the sum of the two individual degrees of freedom. Thus, for Table 1 we had $X^2 = 19.60$ with d.f. = 2. If for the second set of data $X_2^2 = 10.3$ with d.f. = 2; then $X_1^2 + X_2^2$ will have the χ^2 distribution with 4 d.f. provided the second set of data is independent of the first. Hence, for the combined data the critical 0.05 χ^2 is 9.49 to which the sum of the two sample X^2's can be compared. Since, in the case in hand, $X_1^2 + X_2^2 = 29.9$, which is greater than 9.49 we continue to reject the hypothesis of independence.

It is to be noted that in combining X^2's obtained from 2 x 2 contingency tables, the continuity correction should not be applied to the individual X^2's (2).

g. Interpretation of Very Small Values of X^2's. The question sometimes arises as how to interpret a very small X^2. Such a value will arise when the actual frequencies are exceptionally close to those expected on the hypothesis of independence. The problem in this case is not that the actual data are not in accord with expected values, but that they agree too well. The agreement is so good as to arouse suspicion. In such cases, we look for evidence of non-randomness or for conscious or unconscious attempts to manipulate the results. Always, of course, there is the possibility of a clerical error in recording or copying the data.

h. Chi-Square Tests and Analysis of Variance. It is interesting to note the parallel between χ^2 tests and analysis of variance. In general it might be said that χ^2 tests are to attributes data what analysis of variance is to variables data. In an analysis of variance involving a two fold classification, we can test whether the column means vary significantly among themselves, whether the row means vary significantly, and, if there is more than one case per cell, whether there is an interaction between the two classifications. We can make similar tests for attributes data. In Table 1, for example, we can

test whether the proportions of failures of the various types differ significantly among themselves (a χ^2 test of the hypothesis $p_A = p_B = p_C$ with 2 degrees of freedom) or whether the proportion of failures in the bottom block differs significantly from the proportion in the top block (a χ^2 test of the hypothesis $p_t = p_b$ with 1 degree of freedom) or whether there is an "interaction" between position and type of failure (the χ^2 test of independence that is discussed above).

i. Analysis of Individual Degrees of Freedom. Just as in analysis of variance we can subdivide a significant sum of squares into meaningful components having one degree of freedom apiece, so it is possible to subdivide X^2 into meaningful components with one degree of freedom each (2). By this means we can delve deeper into the nature of the dependence that may have been found to exist between two classifications.

Consider the following:

	Classification I			
Classification II	A	B	C	Total
a	20	24	34	78
b	27	25	38	90
c	46	31	26	103
	93	80	98	271

For this contingency table the degrees of freedom are 4 and the value of X^2 is 10.89. This is in excess of the critical 0.05 χ^2 of 9.49 (See table 3), and we would conclude that the classifications are not independent.

With this result we may wish to delve deeper into the causes of the dependence. This might be done as follows: We might first compare categories A and B of Classification I with categories a and b of Classification II in a simple 2 x 2 contingency table. This would yield

Sub-table i

	A	B	Total
a	20	24	44
b	27	25	52
Totals	47	49	96

for which X^2 has the value 0.38. Since the critical 0.05 χ^2 for 1 degree of freedom is 3.84, we would conclude that the dependence in the larger set does not arise from this part of the table.

Next we might take the column totals of the above 2 x 2 table and compare them in a second 2 x 2 table with the frequency in classes

Ac and Bc of the original table. Thus, we would have

Sub-table ii

	A	B	Totals
Column Totals of Sub-table i	47	49	96
c	46	31	77
Totals	93	80	173

for which X^2 is 1.99 which again is less than the critical 0.05 χ^2 of 3.84. We would find no dependence here, although the X^2 is larger than in the first case.

Thirdly, we might take the row totals of subtable i and compare them with the items in cells Ca and Cb of the original table. Thus, we would have

Sub-table iii

	Row totals of Sub-table i	C	Totals
a	44	34	78
b	52	38	90
Totals	96	72	168

For this sub-table X^2 is 0.04 which is again less than the critical value of 3.84. Still we would find no cause for dependence.

Lastly we might set up a 2 x 2 table using the column totals of sub-table iii, the row totals of sub-table ii, and the items in cell Cc. Thus, we would have

Sub-table iv

Column Totals of Sub-table iii	96	72	168
	77	26	103
	173	98	271

For this sub-table X^2 is 9.35 which is much greater than the critical value of 3.84. This would indicate the source of the dependence in the original table. It would be clear that it is the excessively low frequency in the Cc class that is the principal cause of the dependence of the two classifications.

Although the above four component X^2's add up to more than the original total of 10.89, this approximate "subdivision" of the total may nevertheless be a very meaningful procedure for analyzing the intricacies of dependency between the two original classifications (Cf. 2).

j. χ^2 Tests and Regression Analysis. The χ^2 test of independence is usually employed when the categories of the two classifications are qualitative, but it can be equally well employed when one or both of the categories are quantitative. For example, tests of cotton yarn may be classified into intervals of breaking strength on the one hand and intervals of fiber length on the other.

When both classifications are quantitative, it is more usual to test for independence by applying some form of regression analysis. For example, if a significant linear or curvelinear regression is shown to exist, the classifications (in this case variables) are of course dependent and the regression equation becomes an instrument for estimating one variable from the others. Nevertheless, it is to be noted that regression analysis tests for a particular type of independence (say linear independence) whereas the χ^2 test is a test of complete independence.

k. Measurement of Degree of Dependence. In some problems it may be desirable to compute a measure of the degree of dependence of the two classifications when this is found to exist. One of the better of such measures is Tschuprow's "Coefficient of Contingency" (8), the formula for which is

$$T^2 = \frac{X^2/N}{\sqrt{(r-1)(c-1)}} \tag{3}$$

where X^2 is given by formula 1, N is the total number of cases (= 180 in Table 1), r is the number of categories in one classification (= 3 in Table 1) and c is the number of categories in the other classification (= 2 in Table 1). T corresponds to the Pearsonian coefficient of correlation associated with linear regression analysis and affords a measure of comparison between one situation and another.

l. Failure of χ^2 Test to Take Account of Signs. In conclusion it may be noted that the principal disadvantage of the χ^2 test is that in squaring the differences between actual and expected frequencies it fails to take account of signs. It is possible, therefore, for the differences between actual and expected to be all negative for the left hand categories of a given classification and all positive for the right hand categories, without the differences being large enough to give a significant value of X^2. A sign test or a test for runs might in such a case show a significant deviation from the pattern of independence although a χ^2 test would not. When employing a χ^2 test, therefore, the signs of the various deviations should be examined before any conclusion is reached.

A FEW SELECTED REFERENCES

1. William G. Cochran, "The χ^2 Correction for Continuity", Iowa State College Journal of Science, Vol. XVI, No. 4, July, 1942, pp. 421-36.

2. William G. Cochran, "The χ^2 Test of Goodness of Fit", Annals of Mathematical Statistics, Vol. 23, No. 3, 1952, pp. 315-345. This article is a "must" for all who actually employ χ^2 tests. It also has a comprehensive list of references that may be pursued with profit.

3. Besse B. Day, "Application of Statistical Methods to Research and Development in Engineering", Review of the International Statistical Institute, 1949: 3/4, pp. 129-155.

4. Acheson J. Duncan, Quality Control and Industrial Statistics (Homewood, Illinois: Richard D. Irwin, Inc., 1952.)

5. R. A. Fisher and F. Yates, Statistical Tables for Use in Biological, Agricultural and Medical Research, (Edinburgh: Oliver and Boyd, Ltd., 1949)

6. Evelyn Fix, Tables of Non-Central χ^2, (University of California Publications in Statistics, Vol. 1, (1949), No. 2, pp. 15-19)

7. A. Hald, Statistical Theory with Engineering Applications (New York: John Wiley and Sons, Inc., 1952)

8. G. Udney Yule and Maurice G. Kendall, An Introduction to the Theory of Statistics (London: Charles Griffin and Co., Ltd., 1949). A classic in the field of contingency tables.

INDUSTRIAL ENGINEERING AND QUALITY CONTROL

Arnold Longley
Atlas Steels Limited

Although it is customary at annual conventions to review in detail your achievements in the previous year, to pat each other's back, and generally, whilst undoubtedly getting a stimulus from the Association with fellow-members, holding in effect a form of vacation conference, I hope that I shall be able to present some unusual slants that may indicate useful lines of attack in the future for the use of statistical tools.

I would like to divide my talk into four main parts:-

1. Industrial Engineering practices that tend to weaken quality.
2. Quality control practices that could be improved.
3. How a greater reconciliation is possible between Industrial Engineering and Quality Control.
4. Future needs in statistical techniques.

1. INDUSTRIAL ENGINEERING

With an emphasis continuously on greater productivity, a feeling undoubtedly arises in employees that they have to work harder. This is reflected psychologically in less employee attention to quality of product, as well as an actual decline in quantitative performance.

As long as incentive systems exist in industry, ready excuses are available when quality is lowered, and as much of Standards setting belongs in the realm of the intangible, no ready means of demonstrating equity of these standards can exist.

The low caliber of many so-called Industrial Engineers, their lack of capacity, and their corresponding dogmatism, add to quality decline.

The time spent on Methods Simplification by Industrial Engineers is unbalanced in its scope. Too much attention is given to the spectacular, and not enough to quality. Even using some methods shown on Barnes' Motion Study films I have been unable to realize even minor production increases without deterioration in the product.

The Industrial Engineer is inclined to resent interference from Quality Control personnel, and they too, at times, resent suggestions and criticism from other service departments, and this friction being clearly demonstrable to employees, also tends to slackness in inspection personnel.

The conflict between inspection staffs and line supervision receives too little attention from Industrial Engineering personnel; the reason is probably that if any man is doing a really good job, he tends to concentrate on his own part of the work more than anyone else's. Another source of conflict could well be that with two separate groups using mainly statistical methods---Industrial Engineering and Quality Control---there is certain to be a lack of co-ordination unless one is made superior to the other.

Lack of communication often causes even good Industrial Engineers to fail miserably in reaching even foremen with important highlights in technical applications, and a similar lack of communication sometimes exists between the Industrial Engineer and the Quality Control man.

2. QUALITY CONTROL

In common with most other highly technical services in modern industry, the application of the laws of probability and statistical research on industrial quality originated in, and is still largely confined to, the universities. With the king-size mathematical brains that have been, and are being, applied to the problem, it is to be expected that much of the language and arguments used will be unintelligible to many people in industry.

I have had the misfortune to watch two very erudite Professors argue some of the mathematical equations used to demonstrate Quality Control practices in front of a class of more than fifty fourth year University students, without any enlightenment reaching the students---rather confusion.

At the State University of Iowa in 1950 I also watched, at the Management Conference, some excellent expositions by Professor Niles H. Barnard of the University of Nebraska, completely ruined by a long-haired and abstruse dissertation by one of, I believe, the same faculty.

With many foremen in industry unable to grasp a very simple graphical presentation, it is not too easy to get them to see the uses of a technique, that, when presented sounds like Greek to them, and which is apparently, like Einstein's relativity, understandable only by the few.

Other Quality Control practices, with which you are probably only too familiar, and which could be improved are:-

(a) The difficulty in getting girls with lacquered and pointed fingernails to take correct samples of small parts from tote-boxes.

(b) The difficulty in explaining that rolling dice as a means of determining random sampling methods is sensible to the average worker and foreman.

(c) The theory of the normal distribution curve, standard deviations, points of inflexion, "X" and "r" charts, sigma limits and the many other academic applications of probability laws which are difficult to explain to men who had public school education only, particularly when these men have been away from any schooling for twenty years or more.

(d) The fact that a normal distribution curve seldom occurs on most industrial operations but its many variations, multi-peak, triangular and what-have-you do, with the difficulty of assimilation on the part of most foremen to the above, also causes difficulty in understanding.

(e) The feeling of Company management that all material shipped should be exactly according to specification, even though this may never be possible, in most instances creates a reluctance to the acceptance of Quality Control techniques.

Other points could also be added at length. Generally speaking, the simpler the technique is, the easier its acceptance and application become.

Statistical Quality Control is of course not the only type of Quality Control. It is, however, relatively new and a far greater selling job is needed on management and worker before it will begin to enjoy the acceptance it undoubtedly should have. Quality Control as conducted in industry prior to Statistical Quality Control, is still probably the major control in industry to-day.

3. RECONCILIATION BETWEEN INDUSTRIAL ENGINEERING AND QUALITY CONTROL

Many things could be done to assist Quality Control personnel by almost all staff departments, but probably the greatest single contribution to the acceptance of present Quality Control of a statistical nature could be made by doing a better selling job on the Industrial Engineers. With the relatively low caliber of many Industrial Engineers to-day, this would not be too easy to accomplish, but the following methods are suggested for your consideration:-

(a) In large plants an intensive six months course in Quality Control techniques are given to Industrial Engineering personnel, taking such numbers at a time as will enable the Industrial Engineering Department to carry on its present function without any interruption or delay.

(b) Making either Industrial Engineering subordinate to Quality Control, which I am certain most of you here would like to see, or making Quality Control subordinate to Industrial Engineering, which would have immediate acceptance from the Industrial Engineers.

(c) A more concentrated effort on quality in the years ahead on the part of supervisors with possibly some course of a more formal nature to bring the various departments into closer co-ordination. For example, in a steel mill it is not only important that Industrial Engineering and Quality Control be closely knit, but that Metallurgical, Combustion and other services be very closely co-ordinated or integrated.

(d) A greater attempt should be made by both Industrial Engineers and Quality Control personnel, not only to simplify their own techniques, but to describe these in language that the working man can understand, and to remove immediately any techniques which are academic with little practical application. For example, the processing of data to 7 or 8 decimal places, when the original information was accurate to plus or minus 20% only, may produce some very interesting looking tables, the actual value of which is much less than the paper on which they are printed. This remark applies quality equally well to Cost Accounting and similar departments.

(e) It is seldom that the theoretical calculations can be wholly applied theoretically in any field, as, in theory, no allowance is made for the human element. Failure of the human element is responsible for most of the failure not only in Industrial Engineering but in Quality Control and these points should undoubtedly receive more attention than a desire to bulldoze a technique into industry.

In general, the basic factor, that, the more a technique is understood the greater its acceptance and value, should be followed, and it is not customary for academically minded people to be salesmen, and therefore, some explanation from people with salesmanship ability rather than academicians, is undoubtedly necessary in industry to-day.

4. FUTURE NEEDS IN STATISTICAL TECHNIQUES

The type of thinking that is applied to modern industrial Quality Control can be very effectively applied to other, and in some cases prior problems than quality. Some of these are listed below and in the opinion of the speaker should have received considerable attention many years ago:-

(a) Determination of optimum plant sizes. One of the biggest causes of excess cost in industry to-day is the expansion of a company to unreasonable limits, where services such as maintenance, repair, and so on, cost excessive amounts of money due to increased length of transportation and "make-ready" time; so far as I know beyond "playing by ear", little has been done with this most important problem facing industry to-day. We are particularly concerned with this problem at the present time in Canada, as we are apparently entering a period of almost unlimited industrial expansion, where failure to consider this point will certainly keep our living standards lower than would otherwise be the case. This would seem to be an ideal subject for Industrial Engineers and Quality Control personnel, who deal with statistics, to begin on.

(b) Market Research. Although there are many organizations allegedly handling this problem, many statistical applications can be used to give us greater information on this very important subject, particularly in times that are becoming increasingly more highly competitive.

(c) You will probably hear from Mr. Brisley of another statistical application where laws of probability are used to shorten considerably Industrial Engineering time in the use of "the ratio delay study". Although this technique was only introduced into North America by Professor Lee S. Morrow in 1940, it had been used not only by the writer, but by other consulting firms for some 15 years prior to that date in Europe. "The ratio delay study" and its many applications are still in their infancy in North America, and to my mind present sizable cost reduction potentialities to almost all manufacturing industry.

(d) Standards for Maintenance and Repair. One of the most difficult problems industry has to face is the control over maintenance and repair labor and materials. This is too well known to need further comment. One of theapplications using a combination of Quality Control and Industrial Engineering techniques of the laws of probability has proved to

be very effective in determining standards for this highly important group in about a half dozen industries in North America. That is the combination of Gallup Poll techniques plus the collection of statistical data. I could add in passing that repair and maintenance labor and materials in a steel mill accounts for approximately one half of prime cost.

(e) Machine Interference. In the textile and automatic machine industries the table originally prepared by Dale Jones in the calculation of percentages of machine interference based upon man time per machine and number of machines assigned, was extended by myself to cover assignments of 21 to 100 machines for varying man time percentages in 1947, using the formula given by the originator. This has been proven to be very successful in its application to textile plants and companies using automatic screw machines etc. Its use could undoubtedly be further extended in almost all textile industries.

(f) Advertising research. The effects of advertising, by type, through spot surveys and analyses is never felt to be of great importance in non-competitive markets. Simple applications of statistics to advertising results are probably too obvious for me to mention, but I think we shall see an increase in their uses and applications in the period immediately ahead.

(g) In large plants statistical data could easily be collected on tool utilization, machine efficiencies and other subjects where alternative choices exist in machines and type of tool.

(h) Sales. The relative efficiency of salesman, determination of territory sizes, determination of number of customers per salesman, has received some attention in the past. However, the results and the method of surveying these problems has never to my mind appeared very satisfactory, and has certainly not been effective in competitive times. For the past 10 or 12 years most industries have not needed salesmen, but merely travellers, as most manufacturing companies could easily dispose of all they could produce. As times change, however, a great use of both Industrial Engineering and Quality Control techniques could be applied in this field, and would undoubtedly meet with very heated opposition from the sales staffs.

(i) Office Procedures. Probably the largest potential field of improvement in clerical type or salaried positions would be the application of statistical laws and Industrial Engineering surveys to modern office procedures. When it is considered that the ratio, percentagewise, of clerical and administrative help to production workers has almost trebled since 1939 across industry, some idea of the cost reductions that might be achieved is readily available. This will also meet with considerable opposition.

In concluding, if we can eliminate the academic to the minimum degree necessary, if we can co-ordinate Industrial Engineering and Quality Control, if we can apply Quality Control techniques in what appear to be the most profitable and most urgent areas in industry, if we can at all times realize that the duty of a service department is to render service rather than to be authoritative, we shall undoubtedly see an increasing emphasis on Quality Control with a much higher degree of prestige for its personnel than such personnel now enjoy.

ACCEPTANCE INSPECTION FOR THE CONSUMER

Jules Labarthe, Jr.
Mellon Institute of Industrial Research

There is more than a little truth to the observation that the installation of any retail store laboratory for quality control was probably a reflex defensive reaction to the arm-twisting by super salesmen on one side and the shin-kicking by unhappy customers on the other. Even today, some 30 years after the mail-order houses first installed their technical control laboratories, the small group of retail establishments using their own laboratory facilities or employing the services of the many excellent commercial testing laboratories to any great extent still constitute a very small and exclusive group. The ultimate aim of any department store or other retail establishment is to make its goods as attractive as possible to its customers and to cause these customers to come back again and again for their purchases. Another objective is to see that the merchandise stays sold and does not become a customer complaint.

I think in each case it was clearly called to the attention of the laboratory head that this new technical appraisal venture on which the retail establishment was embarking was to serve as the customer's laboratory. In other words, the interest of the store's customer was first the local consumer herself--the one for whom the laboratory was to act. Even without such advice, it would take any store laboratory technical person only a short time to find out that voluntarily or involuntarily he was acting as the appraisal engineer for the ultimate consumers of his community. When a prospective customer first sees and purchases a new kind or brand of merchandise in a store, there are three primary requirements which must be met: first, how does it look; second, how will it wear or serve her needs; and third, how must it be cared for in use. Of these three primary requirements, only the first or appearance angle is generally of significant importance to the store buyer. This, therefore, is the only truly customer point of view element that the buyer will necessarily consider when she makes her selection of goods from the manufacturer. The other two functions to be fulfilled; that is, serviceability and care in use, are really consumer points of view. They are factors that only become important after the article has been purchased and put into service. These two factors are of great concern to the laboratory of the department store, or the mail order house, or the commercial testing establishment whose services have been commissioned by a store. These are the consumer-use or quality control criteria for which the store laboratory is primarily responsible.

Any one of us engaged in this kind of work will freely admit that our results would throw utter confusion into the mind of any statistician who might be charged with the responsibility for determining the statistical validity of our findings. Our sampling must necessarily be very spotty, and all we can hope for is that the samples we are evaluating are truly representative of the manufacturer's average production. Thus, the statistical validity of our results depends entirely upon the quality control of each individual fabric manufacturer or the producer of any other merchandise under examination in the laboratory for its suitability for consumer use; whether it be a

soap, a rug, an electric appliance, or a new kind of clothes pin. In this regard, I am sure that the mail order house laboratories, having closer contact with the producing element of the business, do have a more valid statistical pattern to follow than do many of us in the department store laboratories. Nevertheless, by confining our purchases to the better producers--that is, those of greater reputation for good quality--we are able to protect ourselves to some extent against wide fluctuations in performance.

There are several distinct reasons why a textile product, let us say a dress, a bathing suit, or a negligee, or an article made of some new textile material will be sent to the laboratory for evaluation.

1. The buyer may want to know just what kind of product it is in terms of fiber identity or perhaps she will want to know something about its serviceability in use.

2. There may be a factor of safety involved. Possibly it is something that might be dangerously flammable, or have some other health aspect connected with its use.

3. It might be a prospective private brand article that the store wishes to introduce into its own-brand program.

4. It may be something with which a neighbor down the street is doing a "grand" business and, therefore, the store buyer wishes to get in on this source of profit if it has merit and is not just a "gimmick."

5. It may be an exclusive that has been offered by the manufacturer for a limited period of time, and the store buyer wishes to find out whether it really is a value or a prestige item.

6. Perhaps a customer has complained about the article; therefore, it is necessary to find out whether the customer's complaint is truly a legitimate one or whether she misused the article in service.

The field of textile products, especially garments, is wide open for any kind of technical appraisal on the part of the store laboratory because practically no manufacturers of wearing apparel engage in any kind of technical evaluation of their products. In all fairness to the men's clothing industry, I should probably exempt them from this charge. However, even here, occasionally untested and inferior types of material enter into the construction of a man's suit even in some of the better houses. In women's clothing, however, the statement--sweeping as it may seem--is accurate because here the fabrics are purchased almost entirely because of their appearance and often with no regard for the particular end-use for which a certain fabric was originally dyed and finished. Fabrics for women's wear are bought by the garment maker even without any precise knowledge as to the composition of the fabric itself. The maintenance of quality, therefore, in the field of women's clothing and, to a somewhat lesser extent, in children's clothing is a real challenge to any store laboratory unless rigid specifications have been prepared for each particular end-use item. A new fabric construction or a new color can come on the market from a fabric resource that has devoted a great deal of research and testing to its development, but if it is good it takes only a very short time for copyists, technicians, or dye-houses to come out with a

virtually identical product in appearance and texture. This makes it very difficult for the fabric distributor or for the manufacturer of the garment who wishes to produce the best in his field; the original and the copy look alike; maybe the cost is very slightly different; and not knowing which is the better, he will naturally purchase the cheaper of the two. Maybe the copy is the better, and the manufacturer would be making a very serious mistake to purchase the more expensive one in self-defense.

When the finished garment reaches the store, the origin of the fabric--to say nothing of the dyestuff, the finish, and the original yarn--is lost way back in the complicated chain of textile development. This means that each individual item on which a store buyer wishes to have any kind of technical appraisal must be tested both by chemical and physical means in order to find out how it is going to serve the customer. The buyer will know whether it is appealing to the customer from the standpoint of appearance. It is up to the laboratory and to the quality control program of the store to find out whether it is going to be a proper consumer-use kind of thing to purchase, to promote, and to expect to be a good resale item season after season based on performance and care criteria. This simply means that the store technician must find out how the article will perform in service and what kind of care must be taken of it if it is to stay in service for a satisfactory length of time for the average consumer.

Particularly in the textile field there is a serious dearth of any kind of nationally recognized standards of performance for such end-use products as wearing apparel or household fabric merchandise. This means that each individual laboratory must soon arrive at its own pattern of quality control criteria, its own standards of performance for fabrics entering into various end-use applications and its own end-use performance standards. It was partly because of the lack of such standards that the store laboratories unanimously encouraged and supported the Rayon and Acetate Standards recently promulgated through the American Standards Association under the sponsorship of the National Retail Dry Goods Association.

This standard lists some 31 textile test methods and has set up requirement criteria for 51 different end-use items made of rayon or acetate. The particular standards of performance set up, let us say, for colorfastness to washing, to light, to perspiration, to crocking, to wet rubbing, etc., the shrinkage tolerance on washing, seam strength, the breaking strength of the material, and other factors are based upon background information that we have accumulated through the years both in department store laboratory work and in the evaluation of merchandise by the American Viscose Corporation on whose "Crown Tested" rayon program we depended so heavily in the development of this standard. Extensive use was made also of the experience of other textile producers in the rayon field and of the commercial testing laboratories. These are not ultimate values in terms of maximum quality, but they are values which in our experience do not result in many consumer complaints in service. They can be regarded, therefore, as minimum standards below which there is a danger sign as far as consumer experience is concerned and above which there is a wide open field for a store or for a textile producer to claim plus values over and above the basic standard itself. Now admittedly, very few, if any, store buyers that I know of are as yet conscientiously and consistently ordering their

rayon and acetate garments according to this standard. Nevertheless, it is the tool or measurement eagerly seized upon by these stores' laboratories and these requirements are being set up by the laboratories for any rayon or acetate product submitted by buyers for appraisal. In other words, having this base from which to work we are all of us employing this standard in our evaluation of goods. Perhaps unconsciously we are applying many of these same performance criteria to similar end-use articles made of other textile fibers.

Our tools for quality control are all nationally recognized methods of chemical analysis or physical testing that may be available. Another set of tools, is the previous experience of the laboratory staff in what is or what is not a very successful consumer end-use item as far as actual serviceability data go. This experience is very difficult to pin-point because it depends so much upon the backlog of data accumulated through years of merchandise appraisal, upon the observation of goods that have been returned by unhappy customers, and upon the excellent judgment of skilled buyers and merchandise managers. All of us have had to develop our own tools when there was none available that would answer a particular question we wished to have answered. Thus, we all have a certain number of private test methods which we use and interpret in the light of our own experience. These private test methods were one of the factors that bothered some persons when it came to studying and approving the Rayon and Acetate Standards, and many of these methods of private nature are now being evaluated by such technical societies as the American Society for Testing Materials, dealing particularly with physical testing techniques, and of the American Association of Textile Chemists and Colorists whose domain in the textile research and evaluation field is in the chemical realm. Only recently the American Standards Association explored the venture of preparing similar standards for all textile products used in hospitals, hotels, and other institutions. Admittedly, many of these standards for institutional use will be higher than those required in the average consumer's home. Nevertheless, those of us in department store or mail order house laboratories will watch these standards keenly in order to apply them whenever possible to our own particular end-use specifications. Above all, our participation in the deliberations leading up to these standards before they are finally promulgated will surely make us more critical of our own quality determining methods and in our own quality control program.

One of the new schools of expression in the field of sculpturing has developed what they call "mobiles." These are physical embodiments of the principles of analytical mechanics and consist of arms of varying lengths from which are suspended different weights in the form of grotesque objects or things of incongruous nature, all kept in perfect balance by clever engineering of the whole assembly whether it be supported by a staff rising from the top of a table or suspended from the ceiling. To some extent, our specifications and standards for quality control of consumer end-use items, especially in textiles, resemble these mobiles because if one thing is out of balance it affects everything else. For example, with an existing specification we might say to the producer, "This is all right but you have got to increase the crease resistance of this material." In trying to achieve this by means of adding more textile resin finish, the durability or wear resistance of the fabric may be seriously impaired. In other words, our mobile or standard of quality will collapse. Similarly, we may say to the producer, "The colors are pretty good, but we want more bright-

ness of hue." This can be achieved by use of some different group of dyes or perhaps by adding an excess of color, but the result as far as the sun fastness or the wash fastness of the color or perhaps even its tendency to rub off will seriously affect the happiness of the consumer using the article. In having obtained our brightness which would be a real advantage as far as the initial appearance of the article is concerned, we may ruin it as far as the use or the care corners of the triangle are concerned.

This means that the store laboratory charged with responsibility for quality control, particularly in textiles, must recognize the delicacy of balance between the various elements they are called upon to evaluate in the end-use item. We strive, therefore, to set the actual test requirements severe enough that customers will enjoy the use of the article and that they will not have too great a degree of responsibility or worry about its actual care in service, such as unusual methods by which it must be cleaned, or ironed, or stored. The criteria for the various end-use experiences to which a given fabric may possibly be put will vary greatly. Each of us has worked more or less alone in the past, although we have exchanged some information at meetings of technical societies and in committee deliberations; nevertheless, I would venture to say that our specifications for window curtains whether they be of Orlon, nylon, cotton, or acetate might vary considerably from those of Macy's or Gimbel's, or of the big mail order houses. Our experience would be based upon the actual conditions in Pittsburgh as compared with other cities in which these other stores have the preponderance of their customers. All should agree fairly closely on bathing suits and perhaps on spectator sportswear for women. We would probably not vary greatly in our requirements for men's clothing or for men's shirts or for boy's wear because, in these goods, maximum utility is of primary importance. Nevertheless, we each have our ideas as to the order of importance of various use factors and physical test results, whereas others perhaps would rank them differently. Thus their specification or requirement for quality control could differ from ours. This again is another reason for us to welcome the standards on acetates and rayons because for the first time we have seen in this document some kind of end-use requirement to which all of us can subscribe, and it covers a great many different kinds of end-use items from women's wear to boy's wear to men's wear and even into the field of household fabrics.

Although it is rare for a buyer of a textile product to be concerned generally with the safety of the item in consumer use, this does occur occasionally when there is a potential hazard from fire. We are concerned with the flammability of some of our rather common textile products made of the cellulosic fibers and in certain fabric constructions. It is significant, I think, that the episode of the flaming sweaters of about a year and a half ago did not involve merchandise sold in any department or mail order store. These products were universally obtained through corner textile produce hawkers and salesmen visiting sororities and fraternity houses in some of our college towns. Nevertheless, the episode did have a profound effect upon the various laboratories, and I know in our case at least we had all of the plush and fuzzy textured cotton and rayon toys sent to the laboratory for examination in order to determine whether they were hazardous from flame. Included with safety, I suppose we could legitimately consider the permanence of mothproofing or insect-proofing treatments to various

textile products. We are concerned in this regard not only with the permanence of the finish but also whether the finish is really effective. The problem of skin irritation or of dermititis is so much a personal one that very few laboratories are concerned with it, and it would be virtually impossible for us to evaluate products for this phase of safety in use.

A real part of the consumer's satisfaction with a product that she buys in a store is whether or not it lives up to the representations made for it by the manufacturer and, therefore, up to her expectations. Textile products are rarely over-labeled as far as manufacturers' claims are concerned; nevertheless, a real part of quality control effort does depend upon the verification of the manufacturer's claims for permanence or for effectiveness of a particular finish he may have put into the goods. In the non-textile field, the sin of over-claiming benefits is much more common. Some of these products that are in error do concern textiles because they are detergents or cleaners which are claimed by the manufacturer to be suitable for any kind of textile product. It is not at all uncommon for the conscientious laboratory to reject many of these products from the store shelves even though they may have great merit for certain kinds of application. A case in point is a detergent which we recently rejected; although it was an excellent general utility household cleaner, it was very highly alkaline and had some added free alkali and then when the manufacturer claimed on the label that it was suitable for any kind of painted wall and even suitable for washing of textiles in the home washing machine and that it "would not damage wool", we rejected the product because the consumer would be led to misuse the product based upon what the manufacturer said on his label. Therefore, it is not at all uncommon for the quality control program of a retail establishment to insist upon label changes before a product can be sold. I might say here, too, that some products that are rather highly flammable and which do not carry a caution label cannot be carried in stores having laboratory control until a caution label is inserted or set over the regular label by means of a stamp.

Thus it will be seen that the technologist working in a store laboratory representing the consumer element in the community must be armed not only with the standard test methods and the agreed-upon degrees of interpretation and rating but also must be possessed with a considerable amount of ingenuity as far as improvising appraisal techniques methods is concerned. It must be emphasized that every attempt to evaluate merchandise of the kind I have mentioned today is from the consumer point of view. No amount of pre-evaluation that we can carry out will completely avoid or prevent customer complaints because for every kind of consumer use we can forecase the actual consumer can find many more ways of straining, misusing, miscleaning, and otherwise mistreating merchandise. A very real part of our whole program must always be in the technical evaluation and examination of these customer complaints in order to be sure that the merchandise was serviceable or whether the consumer has in some way mistreated the item. From our own experience based on somewhat over 10,000 individual customer complaints in the textile field, we have come to the conclusion that in roughly 60% of the cases the consumer has misused the articles that have come to our attention. It must be repeated, however, that to approach quality control of department store merchandise from the point of view of customer complaints is a defensive

maneuver taken after the horse has been stolen from the stable. Too often such complaints are made too late to be of real benefit. Occasionally, however, we have found that an article that perhaps soiled very quickly in the hands of the consumer was sent to the laundry or the dry cleaner much earlier than we would normally expect, and the consumer experience has been that the article would not clean satisfactorily. When this happens very early, it is often possible for us to stop the sale of similar merchandise that may still be in stock until such time as we are able to check in the laboratory to find out whether the customer mistreated the article or whether it really was not a cleanable item. I well remember one case when a customer who purchased a suit of acetate had it cleaned and the material faded at the shoulders. This particular cleaning was done only ten days after purchase. The shoulder fading was due to the presence of rubber latex in the shoulder pad, and we found that a small trace of this latex would rapidly fume-fade acetate when various test fabrics of this material were made into soft pillows filled with the shoulder pad material. The manufacturer was advised immediately to substitute another shoulder pad material in his line which was then in production for the full summer season. It is certain that both this manufacturer and his retail store customers, to say nothing of their customers--the ultimate consumer--would have suffered very much greater financial loss had it not been for the more or less accidental rapid cleaning of this particular suit and a prompt and conscientious complaint being made by this individual consumer, followed by prompt action on the part of the store buyer in sending the product to the laboratory for evaluation.

The store laboratory has a real place in the textile quality control picture. It is the only laboratory that is concerned with the entire garment as it is purchased and as it is going to be used by the consumer. A few large textile producers have undergone or undertaken some kind of control similar to this approach using both laboratory test methods and field test techniques, but they are concerned with their own products and, in general, the article that they are field testing is one made more or less to order for them for the purpose. The retail store laboratory gets the run of the mill production, so to speak, and, therefore, our responsibility is much more for the real consumer-experience than it is for the store's own buying guidance. One technique that all of us are interested in and with which we have done very little thus far has been in the field or use-test of finished garments under controlled conditions or by means of completely randomized samples. This kind of evaluation is difficult to control, to interpret, and to finance, but unless it is done we must still regard some of our consumers as being the guinea pigs who are doing the comparison testing of our laboratory results. In other words, perhaps the buyer has sent a product out to the laboratory for complete examination and analysis and testing prior to purchase. There have been times, too, when the sample sent to the laboratory for examination has been a pre-selected quality. I remember one case in which we had an over-stuffed chair which when painstakingly taken apart and all elements of construction had been thoroughly tested and judged the product was approved for sale. A short time afterward, a customer complained about the poor durability of the similar item she had purchased in the store. When the customer's complaint merchandise was examined, the framework was found to be of packing crate quality. Obviously, therefore, the manufacturer's output was not of uniform

quality, and we have since suspected often that their samples might be of superior quality to their ordinary production. This is an example of a product meeting all of the laboratory requirements, and yet when it is in the hands of the consumer something begins to go wrong. These consumers have been the guinea pigs by which certain flaws in the laboratory control picture have been brought out in the open. They may be the ones who are responsible for upsetting the delicate balance of our mobile, our physical structure which we have painstakingly drawn up in a revised specification in order to change or improve some factor which was probably in better balance to begin with.

As spotty as its sample may be and as incomplete as it undoubtedly is, considering the thousand of textile products sold in any store, nevertheless the quality control efforts of department store laboratories, or mail order house laboratories, and of the commercial testing facilities utilized by other stores do play a real part in helping our customers get their money's worth in the products they buy, and when they become the consumer or user of the article they find that their care and use-experience have been simplified for them by what the laboratories have accomplished through specifications and higher standards of quality. The picture is still far from complete and much remains to be done, but I believe that the quality control as far as consumers are concerned rests not a little upon the responsibility, the knowledge, the experience of store laboratory technicians. The store buyers themselves must be made to realize the importance and significance of this to their buying habits. They must be taught to use such things as a Rayon and Acetate Standards and any other similar textile standards which may be forthcoming. The value of a product is not only how much it costs or the effect it has on one's personal appearance, but must always include the length of time it retains this appearance improvement, the degree to which it lives up to the intangible and unknown consumer expectation of the item, and also the way in which it must be cleaned and cared for. The quality control program is an ever expanding one in every store laboratory.

CONTROLLING VISUAL CHARACTERISTICS ON PRECISION PARTS

Aldis S. Hayes
Perfect Circle Corp.

Look in any issue of Industrial Quality Control and you will see many advertisements by companies which have found better ways of measuring things. In many cases, one of the primary objectives has been to minimize the human element. In some, the element of touch or feel is replaced by mechanical devises, while others are so automatic that the inspector himself is eliminated.

These are not isolated cases - they are typical of the unceasing effort to make up for man's inability to do repetative work consistently. There is no such help for visual inspection, because by definition it must include the man.

When we talk about visual inspection and the control of visual characteristics in manufacturing, we are talking about an art. We are not working with precise measuring instruments which themselves can be measured for error, nor are we dealing with specifications which can be tested in a laboratory against the requirements of the product.

We are dealing with the ability of people to judge the attributes of a product, based on, at best, a vague description of what other people think is good for the product. Or stated in another way, it is a problem of dealing with people, what they see, and what they think.

With this in mind, we should consider visual inspection with the idea that improvement - not perfection - is the goal. The returns for attaining a small amount of improvement will exceed those for failing to attain perfection.

Visual inspection has won for itself an unenviable reputation for being the least understood and most poorly controlled of the various operations which take place in a manufacturing plant. This can be attributed to a number of reasons, some of which have been mentioned. However, one of the chief ones is the fact that, unlike other inspection operations, it is usually forced into the position of starting out without a definite objective, or specification. About the most it can expect in the way of help in this respect is some statement such as "Parts must be free of objectionable defects, etc.".

Too often, in self-defense, shop people have decided what kind and degree of attributes are objectionable instead of insisting on having the objectives specified by the engineering department. Thus, the inevitable post-mortem discloses the superior judgement of hindsight. This is not to recommend that shop people discontinue the practice, but rather that they do a more thorough job of it.

The first step in tightening up the operation is to establish standards. To do this, we must decide at what point, as a given type of defect increases in size or number, we will no longer accept a part. Each company must decide for itself, as a matter of policy, who makes this decision, but however it is decided, these are the factors which will influence it. 1. Past practice, 2. Effect on function, 3. Sales appeal.

If past practice is to be the criterion, the reject point can be determined by sampling the outgoing quality or warehouse stock. But to more accurately establish this point, we should be able to measure (no matter how crudely) the defect and classify the pieces into cells to form a frequency distribution. This is desirable because past practice will not be a hard and fast line, and we can better understand it by getting the full picture. It may seem that measuring a characteristic takes it out of the visual classification, but in the experimental work we can afford to do what would not be practical on a production basis.

If effect on function is to be the controlling factor, it can be tested just as would be any measurable characteristic, after the material has been classified as above. The scale for these classes might be length of scratch, depth of scratch, number of pin holes, etc.

When the decision is to be based on sales appeal, the quickest solution is to assemble a suitable sample and test it on the salesmen. It might be possible to present the problem to the customer for an opinion, but this would fall under marketing research and need not be discussed here.

In any case, the final decision must be described as well as possible in the form of a written specification. It should be written in figures not in adjectives. The more explicit this specification can be, the more consistent can be the inspection, for whatever is lacking in the specification must be made up by the inspectors.

After the standard has been selected it should be tested on the people who are concerned with it to see if it is practical from a use standpoint, and really represents the desired quality. It is especially important that sales and engineering people be brought into this because they are often critical of inspection which is less than perfect. Participation will help them understand that visual inspection will fall short of perfection even with the best of inspectors.

Nothing is so useless as a specification which draws too fine a line between different sizes of defects, or which makes too complicated the various combinations which are allowable. For example, in a statement covering foundry defects, it is better to limit the tolerance to one defect of a rather generous size (say .020), than to permit one of .015, two of .012, and three of .010. The final result will be about the same in either case, and the more simple statement will better describe what is actually being done.

The testing can be done by assembling a sample of about 50 pieces which cover a wide range of size and number of defects - but only one type of defect in a test. Some pieces should have no defects, while some should be so obviously unacceptable that no one will pass them. Present the sample to each person once, then several days later a second time, disguising the identity of individual pieces so the second set of decisions will be unbiased by the first.

This will produce as many answers as we have people, and some very interesting results. We are likely to find that someone rejected a piece which we thought had no defect, while someone else accepted one of the pieces which we placed in the sample because no one in his right mind would pass it. Then too, someone (not an inspector) will

almost surely divide the pieces into three piles - pass, reject, and borderline - as though borderline pieces could be set aside and the decision postponed until they either got better or worse. Discounting the obvious radicals though, we will have some useful information here.

With this information we can answer these questions.

1. What is the inspector's ability to make the same decision both times on a given piece? There will always be some reversals, but too many may mean that too fine a line has been drawn between good and bad quality.
2. Does it take too long to judge a piece?
3. Is the dividing line between good and bad a natural division? Quite often this will have to be a compromise between what we would like to have, and what is convenient. For example, we may be willing to accept a 100 micro-inch scratch but decide on 70 because that is just the point at which a scratch can be felt with the point of a dull, soft lead pencil.
4. What is the general acceptability of the standard to engineers and salesmen? That is, is it the proper balance between cost and quality?
5. How much spread is there between engineers and salesmen? Too much difference will lead to grumbling between the extreme individuals.

During this testing, special attention should be given the possibility that this particular characteristic is not objectionable and should therefore not be considered a defect.

One of the most important results of having a standard is the elimination of drift in judgment. This drift may be caused by a shift in process average, because an inspector tends to reject the same percent of pieces in each lot. If he is accustomed to rejecting about 2% and then gets a lot which is 10% defective, he will probably be a little lenient and reject only 8% or 9%. On the other hand, if he gets an exceptionally good lot he will be more critical, figuring that if he doesn't find 2% he must have missed some.

Complaints are another cause of drift. No matter how well an inspection department is organized, receipt of a complaint is the signal to start a tightening-up process. This may go on for several weeks or months, until the increase in cost becomes noticed, at which time a loosening process starts. Like the pendulum of a clock, the only time it is on center is when it is going from one extreme to the other.

It would be impossible to eliminate all of this drift because the human factor of judgment cannot be limited to definite boundaries. Fairly well defined standards will, however, provide some restriction to the drift, and at least enable us to know the point from which we have moved.

No matter how well written and clear-cut the specification might be, it doesn't necessarily follow that the inspectors and foremen will follow it. Months and years of operating on their own initiative will form habits which are not easily changed. To assure conformity it will be necessary to spot check regularly. In fact, it will be necessary

to do this kind of checking as long as we want reliable visual inspection. This will catch any general tendency to lose sight of the objective. It will also spot the occasional individual who feels the right to inspect the way the boss would have told him to inspect if he had been on the ball.

A convenient method of recording this information consists of charting on one graph the percent defective in the outgoing material and the percent good in the rejected pieces. Using the central horizontal line on the chart as zero, plot one above and one below this line. Thus, at a glance one can see if an inspector tends to be too critical, too lenient, or both. If the department operates more than one shift this can be put to advantage by making a chart on each one in order to determine the difference between foremen. The foreman's characteristics should not be overlooked, because he influences his people, and whatever bias he has will very likely show up in the way they work.

Again, this is a problem of human nature, and each person's work reflects the way he thinks. Some people are cautious by nature and will take out insurance by being overly critical. To them, security results from rejecting every questionable piece. Others are more easy-going and can mentally flip a coin on borderline pieces. It is important to recognize these differences in people and make some allowances. If a person is near the mean in his judgment it is better to leave him alone than to point out small errors. Even a mild criticism may produce an over-correction, leaving the situation worse than it was originally. A policy of correcting small errors might very well upset the whole department to the point where no one would be sure just what was expected of him.

With reliable inspection as a foundation, we can with more confidence, consider the problem of control of the manufacturing operations. The sources of visual defects may be: 1. Material, 2. Equipment, 3. Processes, 4. People.

Conventional scientific analysis will determine which of these four is at the bottom of the trouble. A fairly permanent fix can be obtained through the use of receiving inspection, periodic inspection of equipment, experimental testing of processes, etc. - on all but number 4. Nothing so direct and positive will work on people. People respond to persuation, repetition, competition, persistance, and explanations in terms of their own needs.

Incentive is a necessary part of any job - whether it be operating a machine, inspecting, engineering, or supervising. The need is recognized and pretty well taken care of in regard to the quantity produced by production people, but quality incentives are much less common. As with professional and supervisory personnel, too much dependence for incentive is placed on what is called a man's pride in his work - or craftsmanship. There is such a thing, of course, and it has a certain amount of pull, but it is not reasonable to expect it to hold its own against the dollar value placed on quantity.

The unfavorable balance between quality and quantity is further accentuated by another factor. It is the immediate and certain effect which low quantity has on both the man and management, versus the delayed and uncertain effects of poor quality. These things are espe-

cially important in dealing with the control of visual characteristics because the behavior of people is such a large part of the problem.

The problem then is to get the maximum pull with the least delay. These requirements are pretty well met by a quality bonus, or paying for only the good pieces produced. Unfortunately many manufacturing processes make it impossible to accurately assign operator responsibility for defects in every case - a requirement for the two systems just mentioned. The only alternative then is to rely on indirect incentives such as posters, pride in workmanship, and percent defective charts placed in the shop to create competition between individuals and shifts.

In the case of a consistent offender, it is necessary to consider the possibility that the man is not suited to the job. Like Typhoid Mary of medical fame who was an incurable "carrier", some operators are not capable of being alert to the products which they are producing. They have no knack for combining good quality with the press for production. It is not a matter of training or dicipline, it is a matter of instinct. Dismissal may be the answer, but not necessarily so, for many of these people are better than average when working at one of the maintenance trades.

Conclusion: Visual inspection depends almost entirely on the ability of people to see, recognize, and interpret. It is not a science and should not be shied from because it cannot be treated with the conventional scientific methods. Like any other patient, it will respond more to sympathy and enthusiasm than to neglect.

HUMAN RELATION ASPECT OF QUALITY CONTROL

Carl E. Hoover
Perfect Circle Corporation

During the past several years at these conventions, we've been telling one another all about Statistical Quality Control, the things it will do, how to use the techniques and put them to work. We have heard reports and seen exhibits giving evidence of some of the great things that have been done by the use of this tool called Quality Control. These things, together with the tireless planning and promotion of fine conventions such as this, have had much to do with the great and rapid growth of our fine organization, The American Society of Quality Control. However a look around the country will reveal that percentage wise very few of our industries actually engage in practically applying this new tool, especially, so far as manufacturing processes are concerned.

Now, looking in our own back yard, and realizing how long its been since we first had some knowledge of this tool and how to use it, we begin to wonder why we haven't made more progress than we apparently have. It is obvious that Quality Control has not cured all of our ills; in fact in many places it isn't being used as it could and should be used, even by those who profess to having the know-how and recognizing what it will do for them. Why then, aren't we using this tool as well and as thoroughly as we should?

At these conventions and in our own local sections, we talk to one another in the lingo or terminology of Q.C. Back home that kind of language doesn't mean a thing to those who are on a level, parallel with us, much less to the person a way down the line who has to actually put the tool to work. Strangely enough all along the line we find we must deal with people, and in the final analysis, it's people who actually have to put quality control to work on the job, if there is to be any reward for our efforts. People are more and more important in all the things that we do. The days when a foreman could tell an operator to jump and have him ask in return "How high?" are gone forever. People are human beings and as such we had better learn to get along with them and know something about them, and do it quick, if we don't know already.

If we are to get along with people, all people, we must realize that their behavior is predictable. This we find out by watching their reactions to things that we do. It is necessary to be ourselves at all times, and not try to act a part or be someone we aren't. We must purposely over emphasize the other persons point of view. There must be no personal grudges or grievances if we are to gain cooperation and good will, and our plans must be designed to comply with other peoples habit patterns. Patience and tolerance, is a prime requisite. A frank and honest comparison of ourselves with others will remind us of our own limitations as well as our capacities. We must contribute to the welfare of others.

After having had a look at some of the things we must do to get along with people, what about them? What makes them tick? What makes them want to keep on keeping on? Psychologists tell us that all people are motivated by four unconscious desires. The desire for importance, the desire for romance, the desire for life or long life and the desire

for masculinity. The most predominant of these drives is the unconscious desire for importance, and it is inherent in all people. This inherent bud in all of us, in all people, will burst into full bloom with a simple sincere compliment.

Now what do we know concerning the thinking and attitudes of most people; not all, but most people? Well, they worry a lot, not because there is anything to worry about especially, but they still do it. Their feelings are hurt very easily, they are quite deliberate in all the things they do. They greatly resent being ordered to do anything, but are urged to their greatest efforts by praise. They are suspicious of the motives of others and usually are radical in religion and politics. They would rather struggle alone with a problem than to ask for help and would rather work by themselves than with others. Most people prefer books or magazines to big muscle activity of any kind. They daydream a lot, are poor losers, desire fine detailed kinds of work, are moody, and very conscientious.

Misunderstanding causes family dissention, quarreling among friends and neighbors and ultimately leads to war, whether it's between management and labor or between nations, and always, people are involved. Maybe misunderstanding is one of the chief contributors to our apparently slow progress in getting Statistical Quality Control put to work practically. One cannot converse intelligently about nor work wisely with something or anything he does not understand. Jesus Christ, the Master Teacher, was the greatest applied psychologist the world has ever known. He saw the intrinsic worth in every person and knew how to coax that worth from him. The lessons he taught were learned well because he went to the people where they lived and taught them in their own language. The Apostle Paul said, "Unless it is uttered by the tongue speech easy to be understood, how shall it be known what is spoken?" During the last war all JIT instructors were admonished with the phrase that, "if the worker hasn't learned, the instructor hasn't taught." Perhaps as regards this Quality Control tool, we have become so engrossed in our own intelligence and knowledge of it that, we have taken for granted many of the things that may be completely foreign to the worker who must do the job of putting it to work.

What would happen, I wonder, if we would take the time and trouble to tell each and everyone of our workers a little story about Quality Control that might go something like this.

Anything that can be measured or counted follows a definite pattern that has a characteristic shape and it looks like this:

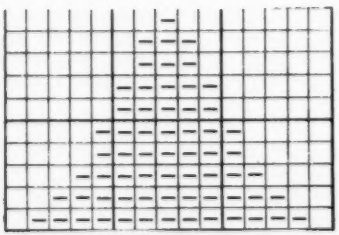


ILLUSTRATION NO. 1

The pattern in Illustration 1 has a striking resemblence to the statistician's normal curve which looks like this:

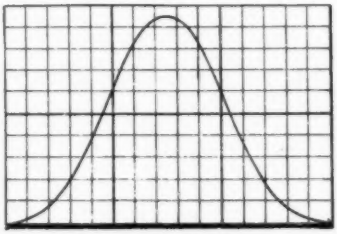


ILLUSTRATION NO. 2

Because of the similarity between the pattern and the normal curve, we are able to use in our industrial operations the peculiarities of the normal curve to tell us what we have done quality wise in our production processes, what we are doing and what we expect to do. Since the quality pattern of a production process can be quite easily obtained and because the peculiarities of the normal curve are known and have been proven to be sound by eminent mathematicians and statisticians, the natural limits within which a process can operate can be foretold, and the number of pieces that can be expected to be found outside specified limits in a large lot can be predicted from a small number of pieces taken from that lot.

As the quality characteristics of individual pieces follow a definite pattern that has a characteristic shape, so do averages of small numbers of pieces follow the same kind of pattern and shape. That is why the Quality Control Chart works. Because the quality characteristics of our manufacturing operations have this definite pattern and shape we have been talking about, and since that pattern is thrown around whatever dimension the process happens to be 'set up' on, it is to the advantage of everyone concerned to have operations 'aimed at' the very center of the limits expected to be held. Operations so directed will offer the fewest border line cases to inspection, and every border line case has an even chance of being misjudged.

In order to get a better picture of the pattern most things follow and have a lot of fun as well, keep score on how long it takes to get to work each day. List down on a piece of paper the spread in minutes the trip may be expected to vary within. Ex: 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18. Note the time when you leave home, check the time again when you arrive at work and place a tally mark after the number denoting the elapsed minutes.

After a couple of dozen trips or so the elapsed time pattern of getting to work will begin to take on an appearance similar to illustration #1. The most usual amount of time it takes to get to work will be the number, opposite which, the greatest amount of tallies appear. Also the shortest and longest times necessary for the trip will be evident by observing the lowest and highest numbers opposite which, tallies fall. Knowing our pattern should help a great deal toward getting to work on time more consistantly. The same kind of score keeping can be

done with your golf game or a crap game and the results are very enlightening.

Think now, of the routine we personally go through in our day to day experience. There is a very definite pattern that has been established by habit. That habit pattern has been formed to provide the most pleasant or the least unpleasant conditions we desire to experience. Whenever something happens that disturbs our habit pattern we are 'upset' even though it may be for the better.

Actually the thing we are talking about is a great deal more than a tool for controlling quality. It is a way of life and follows natural laws.

A machine has a certain amount of variation that is natural to it. This natural variation can be found by making a tally pattern of the quality characteristic being produced, from pieces coming from the machine one after the other and between which no adjustment has been made. The measurement of fifty pieces so taken should form the pattern. The width of the base of the pattern will show the natural variation, and the number opposite which the most tallies fall will show the dimension the machine is 'set up' on, about.

Now if ten machines are producing the same quality characteristic and a tally pattern is made from fifty pieces taken, five pieces from each machine, it will be found that the base of the pattern is about one and three quarters to three times wider than the natural variation of one machine. This is caused by the 'set ups' or adjustments of the individual machines bouncing up and down, high and low.

Without the knowledge of the natural variation of the machine, it is very difficult for the operator to make a wise decision as to what he should do, especially if he is depending upon the measurement of one piece to make up his mind. Always there is one of three decisions to be made; adjust to remove more stock, adjust to remove less stock or leave the machine alone. Since the goal we are striving toward, is to manufacture a product, the quality characteristics of which will distribute themselves uniformly around the centers of the dimensions specified; and have the base of their patterns as narrow as possible, correct decisions and adjustments in the right direction must be made more often.

The control chart furnishes the proper instrument for making correct decisions and indicates the right direction for adjustments to be made. As stated earlier, averages of small numbers follow a definite pattern and shape, but the base of the pattern for averages is narrower than the base of the pattern for individual pieces by an amount depending upon the number of pieces used to figure the average. It is as though the pattern in illustration #1 was turned to the right 90° ; lines drawn horizontally to the right from each end of the pattern and the tally marks scored on a time interval basis. So long as the tally marks are falling within the horizontal lines the indication is that there is no more variation than is to be expected and the pattern isn't being disturbed. When a tally goes beyond the lines in either direction, more variation than should be expected is indicated, the pattern is being disturbed and an adjustment back toward the middle is called for.

In the attainment of the goal we are striving toward, there will be

situations where it is necessary to have the patience to accept some things that can't be changed. The courage to change some things that can be changed and need changing and the wisdom to know the difference. We can at least instruct operators to shoot at a specific dimension instead of an alley between two limits and furnish them with the means of knowing whether or not they are hitting the mark they are shooting at.

Since the Grand Master of the Universe made no two people exactly alike we certainly can expect the things man makes to vary. Variation is the rule, not the exception. Only when the variation gets out of bounds or beyond limits, are we called upon to do something about it. Our responsibility is to know and recognize that a difference, to be a difference, must make a difference and some differences don't, sometimes. Diligent adherence to the dictates of natural laws will lead to uniformity and satisfaction, whether in a manufactured product or in the art of living.

STATISTICAL APPRAISAL OF VACUUM TUBE RELIABILITY

Richard D. Guild
Sylvania Electric Products Inc.

In recent years the term "reliable vacuum tube" has come into widespread use. The need for this type of tube is generally conceded to be primarily military. The failure of a tube in guided missiles or bomber electronic gear may mean an ineffective missile or the destruction of a bomber as well as loss of life. Ineffective missiles and destroyed bombers cost tremendous sums of money. Hence, to the services, the term "reliable vacuum tube" implies that all tubes will operate under use conditions with stable electrical characteristics and no inoperative failures throughout a required time interval.

To meet the needs of the services, the radio tube industry has embarked on a reliable tube program. Realizing that it is impossible to make all tubes perfect, the tube industry has set goals on the rate of inoperative tube failure and stability of electrical characteristics to approach the military need. However, the every-tube-perfect idea can only be approached asymtotically, as the risks inherent in any sampling plan allow some defective tubes to get out the door.

Tube reliability involves three basic concepts:

1. The establishment of the best design for a tube type.
2. Elaborate process control at all points during the manufacture.
3. An adequate acceptance procedure and specification for the finished tubes.

One might question whether these concepts should not also apply to ordinary commercial tubes. Obviously, they do; however, economics must be considered in determining the degree to which the concepts are carried out. In comparing the two grades of tubes, more cost can be incurred to fulfill the three concepts in reliable tube manufacture.

Sylvania's approach to the third concept, the specification and acceptance procedure for the reliable tube program, is as follows:

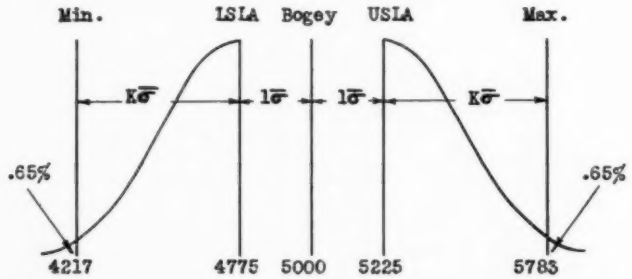
The first step required to appraise tube reliability is a specification in which the various testing conditions and specification limits are defined. To set proper testing conditions and specification limits, the application of a tube type has to be reviewed with the users and the process capabilities under various testing conditions have to be explored. Usually the application is the most important factor in setting specification limits; however, it is important to be aware of process capabilities before specification limits are established. Setting specification limits for characteristics distributed normally or non-normally will be further discussed later in conjunction with acceptance procedure.

The general intent of the lot acceptance procedure is to accept initially each electrical characteristic individually, either by an attributes plan, or the combination of a variables plan and attributes

plan. In this way important electrical characteristics may be controlled by tight operating characteristics and other less important electrical characteristics may be controlled by less severe operating characteristics. Electrical characteristics subject to acceptance by both variables and attributes are controlled not only for Minimum and Maximum, but also for production lot mean and standard deviation, thereby imposing strict requirements for the shape of the lot distribution inside the specification limits for individual tubes. After production lot acceptance has been achieved on the individual characteristics, attributes plans covering groups of electrical characteristics are employed to control total defects.

To elaborate on establishing specification limits and acceptance procedure, let us consider the electrical characteristic transconductance Sm on a typical reliable subminiature. This electrical characteristic is of critical importance for many vacuum tube types. An Sm rating of a typical reliable subminiature is 5000 umhos. This figure 5000 is normally called design center or bogey for the characteristic. To set a Minimum and Maximum specification on individual tubes, daily samples of 24 are read for a 20-day period. It is assumed that prior to the 20 days the design and processing of the tube type has been settled and that the ensuing 20 days should exhibit a controlled process. The mean $\bar{X}$ and the best estimate of population standard deviation σ for each group of 24 is computed, using the range method to compute standard deviation. A pattern of $\bar{X}$ and σ values is plotted and examined for trend or change in level. Assuming both statistics show control, the process average $\bar{\bar{X}}$ and the mean process standard deviation $\bar{\sigma}$ are computed. A desired condition is to find that the process average is located at bogey.

The value of $\bar{\sigma}$ is utilized in establishing limiting values of population mean called LSLA and USLA. LSLA and USLA are placed arbitrarily, however, with hindsight, at $1\bar{\sigma}$ below and above bogey respectively. Maximum and Minimum values are placed K multiples of $\bar{\sigma}$ from the USLA and LSLA respectively, K representing the normal deviate and chosen so that the AQL value for the electrical item is included in each tail. Pictorially, assuming $\bar{\sigma}$ = 225 in the transconductance example, the above looks as follows:



AQL = .65%; K = 2.48
Min. = LSLA - K$\bar{\sigma}$
Min. = 4775 - 2.48 (225)
Min. = 4217

AQL = .65%; K = 2.48
Max. = USLA + K$\bar{\sigma}$
Max. = 5225 + (2.48)(225)
Max. = 5783

Rounding off the Maximum and Minimum values computed statistically to more appropriate specification values, Maximum is set at 5800 and Minimum at 4200. Acceptance of these values as Maximum and Minimum would depend to some extent upon tube application; however, the values as established are considered indicative of process capabilities. The arbitrary choice of $1\overline{\sigma}$ for locating the limiting values of population mean has been developed by experience. Lot by lot pattern of mean and standard deviation over a two-year span on particular types shows process levels within the $\pm 1\overline{\sigma}$ from bogey to be compatible with practical operation.

The above procedure is used to set Minimum and Maximum on all important characteristics that distribute themselves normally. With Minimum and Maximum values established, attributes testing in general is administered by using plans from Mil.Std.105A. The basic philosophy for using Mil.Std.105A is the provision of large sample sizes which have operating characteristic curves with high discriminating power. Several electrical characteristics such as filament current, screen current, and plate resistance are subject only to control by attributes.

For some characteristics, such as transconductance, attributes control is not considered sufficient. Here, control on the mean and standard deviation is also provided to control the distribution shape of the final product inside of the specification limits for Minimum and Maximum. The standard deviation to be controlled is defined as the Maximum rated standard deviation MRSD. MRSD is equated to 1.33 times the mean process standard deviation $\overline{\sigma}$. The control limit on MRSD is called ALD or acceptance limit for sample dispersion. ALD is computed so that the probability of acceptance is 95% for a production lot whose standard deviation is MRSD.

$$ALD = MRSD + 1.645 \frac{MRSD}{\sqrt{2N}}$$

when N = sample size

The control limits on the mean are called UAL and LAL or the upper and lower acceptance limits for sample mean. LAL and UAL are computed so that the probability of acceptance is 95% for a production lot whose mean is equal to LSLA or USLA.

$$LAL = LSLA - 1.645 \frac{MRSD}{\sqrt{N}}$$

$$UAL = USLA + 1.645 \frac{MRSD}{\sqrt{N}}$$

when N = sample size

The constant 1.33 above has been selected to insure that a production lot made up of sub lots whose means extend from LSLA to USLA will not be rejected for standard deviation unless the average standard deviation of the sub lots exceeds $\overline{\sigma}$. It is realized, however, that a production lot with mean located at USLA or LSLA could be accepted with a standard deviation MRSD but would violate the specification requirements for Minimum or Maximum. In such instances an attributes plan would likely reject the lot. The constant is realistic but somewhat arbitrary as it is dependent upon the number and position of sub lot means located with the interval $\pm 1\overline{\sigma}$ from bogey that may be represented in a production lot or the number of sub lot means within the

interval $\pm 1\bar{\sigma}$ that may be represented in a warehouse full of production lots. A detailed mathematical treatise supporting the choice of this value is covered by an unpublished paper--"Maximum Rated Standard Deviation"--dated September 5, 1952, by D. R. Meierdiercks.

Because of the variables control, one might think that attributes inspection is an unnecessary item. One reason has been pointed out why the attributes control is also a necessity. Another reason for the necessity of supplementary attributes control is that an occasional lot will have an unsymmetrical distribution form resulting from assignable causes.

Characteristics which have inherently non-normal distributions are inspected only by attributes plans taken in general from Mil.Std.105A. Most of these characteristics have J-shaped distributions. Specification limits are established on J-shaped characteristics by evaluating composite data taken over a period of time. A limit is arbitrarily selected which has a percentage of individual readings outside. This percentage is usually near the AQL value desired on the characteristics. Tube application is also of major importance in setting these limits.

The question of equipment accuracy is sometimes advanced as a deterrent in administering an acceptance procedure. Certain characteristics are 100% production tested to the Maximum and Minimum limits; hence, AQL values on these items serve as a compensating effect for equipment error. Characteristics that cannot be production tested have to have an AQL large enough to include equipment error. Equipment error is unimportant on some characteristics which do not require narrow tolerance limits in application.

Consequently, a general picture of the initial acceptance procedure finds normal and non-normal electrical characteristics controlled individually by attributes. Certain critical characteristics simultaneously use variables control with attributes control.

Lots rejected by variables control constitute a problem to the producer. Sometimes reprocessing is the answer; however, if this is not feasible salvage of good tubes becomes the problem. Obviously, merely re-screening the lots to Maximum and Minimum will not make them acceptable. Re-screening, however, to calculated limits inside Maximum and Minimum will accomplish this job. A paper on truncation of rejected lots is to be published in the near future by G. V. Herrold of Sylvania.

Even though vacuum tubes meet initial acceptance procedure they are expected to give serviceable performance through a time interval. Hence, life testing vacuum tubes is as important as initial acceptance procedure. Commercial tube life tests are accepted by a procedure under which defects of critical importance are not differentiated from the less important defects. Adding more types of defects to a life testing specification can be accomplished usually at the expense of loosening acceptance criteria and at the same time giving less importance to the more critical defects. Consequently, it became mandatory that a reliable tube life testing specification be developed to overcome these difficulties.

The general intent of the lot acceptance procedure on life is to

have a tight control on inoperatives and transconductance slump with somewhat looser controls on other characteristics. End-point specification limits for the various characteristics on life were determined in general in accordance with the previously mentioned methods.

To elaborate on the life testing procedure, a 150-tube sample, all tubes meeting initial specification (Mil.Std.105A lot size 1301--3200), is carried through 100 hours life and a 20-tube sample selected from the 150-tube sample is life tested 400 additional hours. The 20-tube sample is earmarked for the 500-hour test at the beginning of the life test. The first inspection takes place at 1 hour on ΔSm for the individual tubes in accordance with Mil.Std.105A plan at 10% AQL. Rejection at this point disposes of the lot for reprocessing.

The second inspection takes place at 100 hours where the 150-tube sample is required to meet a .4% AQL on inoperatives and a 1.0% AQL on a transconductance end-point. Because of the large sample size at this point, the high discriminating power of the OC curve gives excellent protection against allowing bad lots for these two items to pass. Hence, acceptance at this point gives sufficient information to continue life testing on a 20-tube life sample. Final acceptance is based on the 20-tube sample as of 500 hours life (the 500 hours includes 100 hours as part of the 150-tube sample). At 500 hours inoperatives and electrical characteristic failures are accepted by individual acceptance numbers and a combined acceptance number for total defective tubes. Failure to pass the individual criteria or the combined criteria will cause lot rejection. A lot may also be rejected at a 300-hour reading interval if the 500-hour requirements are violated at this point. Again at these two points the controls for the various items are relative to the importance of the item, inoperatives receiving the most stress. The end-point characteristic section of a typical intermittent life test specification, type 5840, is shown in the chart accompanying this paper.

The life test specification shows the test, test conditions, inspection level (Mil.Std.105A), AQL, and Minimum and Maximum limits. The points of interest on the specification are the .4 AQL on inoperatives at 100 hours, the control of ΔSm at 1 hour and the list of acceptance numbers on the various characteristics at 500 hours. It might also be pointed out that Note 18 referred to on the 300 and 500 tests states that no more than the 4 defective tubes are allowed on the 300 or 500 hour test.

To conclude, an attempt has been made to establish realistic specification limits compatible with process capabilities and tube application. Attributes testing initially and on life on all important items has been made by sampling plans with high discriminating power. Major characteristics initially have controls on mean and standard deviation that control the shape of the distribution of final product within specification limits. The life test has been established to give a high degree of protection against passing lots having excessive inoperative failure rate and to control major electrical characteristics individually. Although the every-tube-perfect goal cannot be literally achieved, it is still being approached.

END-POINT CHARACTERISTIC SECTION

TYPE 5840 LIFE SPECIFICATION

CHARACTERISTICS (Note 17) Intermittent Life Test(1)End-Points	Conditions	Unit	Insp. Level	AQL	Allowable Number of Defects Per Char.	Min.	Max.
One Hour Test--Sample A							
Transconductance change of individual tubes from initial	E_f = 6.3V	$\Delta Sm(1)$ %	Note 20	10.0		--	8.0
100 Hour Tests--Sample A							
Transconductance (1)		Sm(1) umho	Note 20	1.0		3250	--
Inoperatives (Note 19)			II	0.4			
300 and 500 Hour Tests Sample B (Note 18)							
Inoperatives (Note 19)					1		
Heater Current		I_f mA			4	135	165
Heater-Cathode Leakage (1)	E_{hk} = +100Vdc	$+I_{hk}$ uAdc			2	--	10
Heater-Cathode Leakage (2)	E_{hk} = -100Vdc	$-I_{hk}$ uAdc			2	--	10
Grid Current (1)		$I_c(1)$ uAdc			2	--	-0.3
Grid Current (2)	E_f = 7.0V	$I_c(2)$ uAdc			2	--	-1.0
Transconductance change of individual tubes from initial	E_f = 6.3V	$\Delta Sm(1)$ %			1	--	25
Transconductance change of individual tubes from E_f = 6.3 to E_f = 5.7		$\Delta Sm(2)$ %			4	--	15
Electrode Insulation,g-all		Rg-all Meg			4	50	--
Electrode Insulation,p-all		Rp-all Meg			4	50	--

ACCEPTANCE INSPECTION BY VARIABLES, CONCEPTS AND PROCEDURE FOR

Harry M. Rosenblatt
Quality Control Division
Bureau of Ordnance
Department of the Navy

Abstract:

Some remarks on acceptance inspection and a comparison of attributes and variables plans are made in the Introduction. Various forms of acceptance inspection by variables including single and double sampling plans, known and unknown standard deviation plans and plans for single and double specification limit test are presented in Section II. In Section III a Standard set of variables plans, Naval Ordnance Standard 80 is introduced and compared with MilStd 105. Also, remarks on the standard deviation and the range are given in this section. Finally a mathematical discussion of known and unknown standard deviation plans for single limit test and single sampling is given in the Appendix.

I INTRODUCTION:

Acceptance inspection may be considered a procedure for making decisions; decisions, after inspection, as to whether to accept or reject a quantity of product submitted by a supplier. Statistical sampling plans have been made a vital part of the procedures because it is the effective tool for providing in advance a desired control over the correctness of these decisions. Dependence on 100% inspection cannot be accepted by directors of acceptance inspection programs not only because it is impractical and costly but because experience has shown it is seldom 100% effective in locating defects, provides no predictable control over decisions, and leads one to a false sense of security. The control over decisions obtained by a statistical sampling plan is shown by the operating characteristic curve of the plan. If a lot of any given percent defective is submitted for acceptance, the operating characteristic curve shows the probability of deciding to accept the lot. This probability multiplied by 100 gives the percent of decisions which will be made in favor of acceptance.

In the Bureau of Ordnance acceptance inspection instructions identify the unit of inspection, list and classify the defects which define a defective unit, specify acceptable quality levels (AQL's) in terms of percent defective and provide a reference to the applicable sampling plans. Most often inspection by attributes is appropriate and MilStd 105A "Sampling Procedures and Tables for Inspection by Attributes" provides an ample variety of sampling plans from which the inspector may select, according to definite rules, the sampling plan required for his inspection assignment.

Inspection by attributes has certain definite advantages. The two most important advantages are relative ease of inspection and simplicity of application of the sampling plan. Each unit in the sample selected is classified as defective or non-defective in accordance with whether there is one or more defects on the unit. The total number of defective units in the sample is counted and this total is compared with an acceptance and rejection number in deciding whether the lot from which the sample was selected shall be accepted or rejected. A third important advantage of inspection by attributes is that the protection

afforded by one attribute sampling plan can apply to one defect or to any number of defects for which inspection is required. Of course, the more ways there are for a unit to be classified as defective the more stringent is the requirement on each defect if the total number of defectives is not to reach the rejection number. A fourth advantage is flexibility of application to maintain desired protection of the sampling plan when the total number of defects in a defect class are split into separate groups for inspection at separate stations. In this case if the sampling plan intended for the class is applied to each group separately and defects are independent there is considerable loss in the protection afforded by the sampling plan over all the defects. The protection decreases the greater the number of separate defect groups. Nevertheless, in inspection by attributes the protection of the sampling plan over all the defects is closely maintained simply by accumulating the number of defectives found in all groups and comparing the total with the acceptance and rejection numbers.

This shows the adaptability of the attribute sampling plans in applications over more than one defect. In contrast, variables sampling plans which are presently available are applicable to a single quality characteristic and thus only provides protection over this one defect. The advantages listed above are usually the factors in favoring inspection by attributes. In addition, where attributes inspection is favored, total sample size, although larger than that required for variables inspection, is usually not a major consideration and measurement data even if obtainable on a particular defect is of no real interest.

However, there are instances where the total sample size required for testing a particular characteristic by attributes is of major concern, either because the test is destructive and the product expensive or for other reasons the cost of test facilities for obtaining an observation is very high. Usually in such instances it is desirable to obtain as high a degree of assurance as possible as to the correctness of acceptance decisions, but yet hold the sample size to a minimum. If the test of the quality characteristics can provide measurements on a continuous scale and if these measurements are approximately normally distributed, an acceptance inspection procedure can be devised which will give the same protection afforded by the attributes sampling plan but at considerable reduction in sample size. The measurements are often referred to as variables data and the acceptance inspection procedure inspection by variables.

Usually in these cases,because of the importance of the product, there is also a need for measurement data for other purposes than for acceptance inspection. Some of the other uses for these measurement data are:

(a) In review of specification limits to aid design engineers in setting realistic limits, or in advising them of the maximum values to be expected for their consideration in revising designs;

(b) In control chart studies to determine homogeneity of a production process;

(c) In analyses comparing differences due to product components, and manufacturing or test conditions, as an aid in tracing cause for product quality;

(d) In statistical studies to obtain best estimates of the variability of the measurements, that is, of the standard deviation, and where possible to utilize this information to obtain still further reduction in sample size than that already obtained in variables inspection where the standard deviation is unknown.

While the reduction in total sample size using inspection by variables is often considerable, it is not obtained without some payment of another kind. This is payment of an administrative nature in the form of possible additional cost to make a single measurement and of increased difficulty in applying the acceptance criteria of the sampling plan. Instead of merely counting the number of defectives in the sample as in the case of attributes inspection, in inspection by variables one must be willing to perform some calculations on the sample measurement data. Fortunately the same calculations which are required for acceptance inspection can also be used in any of the studies mentioned above. In addition, if the data and calculations are presented on a form in a systematic manner much saving in time and effort can result in utilizing the acceptance inspection data for these studies.

The calculations on the sample data are to estimate the average (μ) and the standard deviation (σ) if it is not known, of all units in the lot for the particular quality characteristic being tested. The arithmetic mean ($\bar{x}$) of the sample measurements is an indication of the central tendency of the measurements; it is the statistic used to estimate the lot mean (average). The standard deviation(S) of the sample measurements indicates the amount of spread of the measurements about the mean; it is the statistic used to estimate the lot standard deviation.

The arithmetic mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

where the x_i, $i = 1, 2, \cdots n$ are the individual n observations.

The square of the standard deviation is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

For computational purposes S^2 can be written as

$$s^2 = \frac{x_1^2 + x_2^2 + \cdots x_n^2 - (x_1 + x_2 + \cdots x_n)^2/n}{n-1}$$

The standard deviation is then

$$s = \sqrt{s^2}$$

II TYPES OF VARIABLES SAMPLING PLANS:

The sampling plans for inspection by variables discussed here are those which are analogous to attribute sampling plans in that they pertain to the measurement of quality in terms of percent defective. Other plans which are outlined in the literature concern the control of the average, or of the standard deviation. For control of the percent defective several types of plans are now available, each to meet a different requirement. These are plans for single or double sampling, single or double specification limit test, and known or unknown standard deviations. Double sampling plans and proper procedures for double

specifications limit tests have been developed at Stanford University and are presented by Bowker and Goode in their book "Sampling Inspection by Variables". An excellent presentation of single sampling plan and a discussion of sequential plans by variables is given by W. Allen Wallis in "Selected Techniques for Statistical Analysis" by Columbia Research Group. A comparison of sample sizes required for single and double sampling by variables with unknown standard deviation and for single sampling by variables with known standard deviation with single and multiple sampling by attributes is given in Table A.

Single Specification Limit Test

The most frequent case of inspection by variables is that of single sampling and unknown standard deviation. For a single specification limit the acceptance criteria is usually written as follows:

If measurements which exceed an upper specification limit U represent defectives, then a decision to accept the lot is made if $\bar{x} + ks \leq U$

If measurements which are less than a lower specification limit L represent defectives, then a decision to accept the lot is made if $\bar{x} - ks \geq L$

Double Specification Limit

If a double specification limit is imposed, that is, a unit is defective if its measurement exceeds the upper limit U or is less than the lower limit L, the acceptance criteria is slightly different than for a single specification limit test. In this case it is desired to control decisions with respect to the total percent defective regardless of whether all defective units are at the upper specification limit U, or the lower limit L, or divided in some fashion between the two limits. If the defectives are only at one limit then two single limit plans can be used satisfactorily. In this case, the protection of both plans is that given by the one plan corresponding to the limit at which defectives occur. However, where the defectives are divided between the limits there is loss in protection if two single specification limit sampling plans are used for a double specification limit test. The loss in protection, however, is largely regained if a maximum allowable standard deviation is imposed in addition to the requirements of the two single limit sampling plans. The situation is illustrated in figure 1. The triangle in figure 1(a) shows the acceptance region for two single limit sampling plans. The proper acceptance region lies below the curve in figure 1(b). A graph of the acceptance region is given to the inspector who plots his computed values $\bar{X}$ and S. If the point falls below the curve the lot is accepted, otherwise it is rejected. The area above the curve but within the triangle in the region for

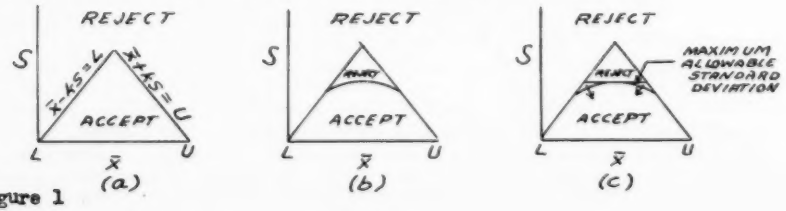


Figure 1

possible values of $\bar{X}$ and S when the lot should be rejected for a double specification limit test, but would be accepted improperly if two single limit plans were used. The trapazoid in figure 1(c) forms the acceptance region by imposing a maximum allowable standard deviation (MASD). It is simple to apply and compensates for much of the loss in protection when two single limit plans are used. The MASD is computed so that it is tangent to the highest point on the curve which bounds the true acceptance region. There are still some lots which may be accepted improperly. These will be the few lots whose values of $\bar{X}$ and S may fall between the curve and the MASD. If a MASD is used the acceptance criteria for a double specification limit test can be written as follows:

Accept the lot if the following three conditions hold:

1 $\bar{x} + ks \leq U$

2 $\bar{x} - ks \geq L$

3 $S \leq MASD$

If any one condition fails, reject the lot.

Double Sampling

In double sampling the procedure is as follows when we assume we are dealing with an upper limit U:

Select a first sample of size n_1. Compute $\bar{x}_1$, and s_1, for this sample.

Accept if $\bar{x}_1 + k_a s_1 \leq U$

Reject if $\bar{x}_1 + k_r s_1 > U$

If the lot can neither be accepted nor rejected select a second sample of size n_2.

For the combined sample of $n_1 + n_2$ items compute $\bar{X}$ and S.

Accept if $\bar{x} + k_t s \leq U$

Reject if $\bar{x} + k_t s > U$.

As in attribute inspection double sampling plans usually result in a smaller average sample size than for corresponding single sampling plans, particularly if the quality of submitted product is very good or very poor. The second sample, however, when required, adds to computations required for applying the plans.

Known Standard Deviation Plans

When the standard deviation, σ, is known the required sample size is much less than for a corresponding variables sampling plan for unknown standard deviation. Using this smaller sample size the acceptance criteria can be applied the same way as for plans with unknown standard deviation using the same constant k, but replacing s by the known value of σ. Often in a process resulting in variables data the standard

deviation remains in control even though there are large fluctuations in averages. Thus during the period of applying plans for unknown standard deviation, the resulting data should be maintained in form for analysis and studies should be made to determine if the standard deviation is sufficiently stable for the application of known standard deviation sampling plans. If this can be done considerable saving in sample size will be the reward. In addition computations will be simplified since the acceptance criteria will be of the form

$$\bar{x} + k\sigma \leq U \quad \text{or} \quad \bar{x} \leq (U - k\sigma)$$

and only the arithmetic mean need be computed.

The Assumption of a Normal Distribution

The constant k and the sample size n are determined based on normal distribution theory so that the desired control over decisions as to the acceptance of lots of varying quality (i.e. percent defective) is maintained. The accuracy of the control of decisions for a given percent defective depends on the extent to which the distribution of the quality characteristic deviates from a normal distribution. Even if there are deviations from normality consistent decisions with respect to lot acceptance will still be made, however, the percent defective for which these decisions are made will not be as intended.

A sufficient approximation to normality in order to closely maintain the intended relation between percent defective and decisions to accept or reject often is not difficult to be satisfied unless control of decisions is required at very small percents defective, say less than 0.1%. For accuracy of decisions when the percent defective is very small, close agreement with a normal distribution is required at the extreme tails of the distribution. This close agreement at the tails is seldom realized even though the observed distributions may be a satisfactory approximation otherwise.

Where the deviation from normality is extreme this fact should be sufficient reason for not using variables sampling plan for acceptance and rejection. In such cases if savings in total sample size is still a prime consideration and measurement data are desirable a combination of variables and attributes inspection can be used.

III NAVAL ORDNANCE STANDARD 80 - A STANDARD FOR INSPECTION BY VARIABLES.

Introduction:

In the Bureau of Ordnance there was a real need for a standard set of sampling plans for inspection by variables which could be released to the Naval Inspection Service and to Naval Establishments. Naval Ordnance Standard 80 "Sampling Procedures and Tables for Inspection by Variables" was prepared for this purpose. The Standard was also prepared in order to eliminate the need to provide a tailor-made plan whenever inspection by variables was desirable. These tailor-made plans previously had to be written into the specifications and other inspection instructions. Preparation of the Standard makes it possible to provide sampling plans for inspection by variables simply by referencing OSTD 80. The objective to provide a standard is in line with the Bureau of Ordnance policy to have uniformity of inspection by all inspectors for all contractors on the basis of a

standard set of sampling plans which can be applied in a uniform manner. For Bureau of Ordnance purposes, it was important to have the standard for variables inspection similar to MilStd 105A in format and procedure in order to minimize the learning process by inspectors already familiar with the attributes document. Where appropriate the text has been taken directly from MilStd 105A. OSTD 80 now serves as an alternate to MilStd 105A in the cases previously mentioned where measurement data are desirable and savings in total sample size is of utmost importance.

It is not intended that these variables sampling plans be used indiscriminately simply because it is possible to obtain measurement data. Because of the added administrative difficulties of applying the plans, indiscriminate use to cases which are not clearly beneficial will be detrimental to the effectiveness of variables inspection as a procedure by engendering resistance to it.

The reduction in sample size and further use of the data should clearly offset the possible required increase in cost of gaging equipment to obtain measurements, extra computations and handling of the data.

As in all variables sampling plans, the basic assumption made is that the data represents a sample from a population which is normally distributed. It is still, of course, possible to use the procedures of OSTD 80 when this is not true, but the extent of the deviation from normality determines the reliability and correctness of the decisions made. Usually the deviations from exact normality are not of a serious enough nature to prohibit the use of OSTD 80. Where a normal distribution is truncated, and where the government desires a combination of mixed variables - attributes inspection is used.

The standard deviation rather than the average range is used as a measure of dispersion of measurements in applying the procedures of the Standard. A discussion of this point is given later on.

The plans in OSTD 80 judge quality by means of percent defective above or below a specification limit. They are single sampling plans, developed for a single specification limit test and unknown standard deviation.

For the double specification limit test, OSTD 80 provides for the use of two single sampling plans, one for each limit separately and imposes a maximum allowable standard deviation (MASD) for cases where defective units may fall outside both specification limits simultaneously. If any one of the three criteria are not met the lot is rejected.

As in MilStd 105A the variables sampling plans incorporate a decreasing producers risk with increasing lot size. This feature, the lot tolerance percent defective for each AQL class, and sample size, is shown in a series of tables appended which are extracted from OSTD 80. Other features of the Standard for selection of Inspection level, sampling plans, and severity constants for normal, tightened and reduced inspection are illustrated by these tables.

In all there are 39 sampling plans, 6 different AQL's and 7 different sample sizes. There are three levels of inspection for normal and tightened and one level for reduced inspection.

Criteria in deciding lot acceptance is provided by computing a lot quality index from the sample measurements and comparing this index with an acceptance constant. The lot quality indices are computed in a consistent manner for an upper or a lower specification limit.

$$\text{Quality Index} = \frac{\text{Specification limit - Average}}{\text{Standard Deviation}} + 10$$

Aside from computational consistency, the expression for a quality index results in a numerical distinction between quality indices for each limit. For an upper specification limit the greater the quality index the better the quality; for a lower specification limit the smaller the quality index the better the quality. The numerical distinction is intended to draw attention to the quality percent defective at each specification limit separately for action purposes, and also to permit easy association of a quality index with the appropriate specification limit. Also introduced are severity constants which are compared with process quality indices computed over a series of lots to consider a change from normal, to tightened or reduced inspection.

The quality index has also been introduced to provide a meaningful quantity to the inspector which he can interpret in the same way regardless of the quality characteristic inspected or its unit of measurement. The acceptance constants and severity constants can likewise be considered quality indices representing standards for lot acceptance or change in severity of inspection, respectively. These quality indices are simply normal distribution deviates, negative or positive, increased by ten. Thus for the case of a normal distribution there is a simple conversion from quality index to percent defective. The direct conversion is provided in Table V of OSTD 80.

The operating characteristic curves which defines the protection afforded by each plan in making correct acceptance decisions are included in the Standard. The OC curves should be referred to as an aid in the determination of what plan to select on an individual basis by comparing plans as to their relative cost and protection.

Many of the operating characteristic (O.C.) curves match closely those of MilStd 105A, so that it is clear from this, that the use of variables inspection where appropriate, provides the same protection for smaller sample sizes.

Table B gives a comparison between MilStd 105A and OSTD 80 on the basis of sample size and the LTPD (lot tolerance percent defective). The LTPD gives the quality of the lot which has a 10% probability of being accepted.

It may also be noted that no involved mathematical formulae appear in OSTD 80. All the steps in the application of the procedures have been reduced to a form to be filled in by the inspector.

Procedure for Use of OSTD 80 and Comparison with MilStd 105A

Paragraph 14 of OSTD 80 provides the inspector with a list of all the tables in the Standard and explains their meaning and use for easy reference in applying the plans. Perhaps the simplest way to follow the procedure for the use of OSTD 80 is to compare it with MilStd 105A with which most of us are familiar.

	Inspection by Attributes MilStd 105A	Inspection by Variables OSTD 80
1. Given: Lot size, Inspection level	From Table III obtain sample size code letter.	From Table III obtain sample size code letter.
2. Given: The (AQL) Acceptable Quality Level	From Table IV select sample size and find acceptance and rejection numbers, a, and r, respectively	From Table IV select sample size and find acceptance constant, A_U or A_L for Upper or Lower specification limit, respectively.
3. Inspector then procedes to:	Select sample, inspect for full list of defects and count the number of defectives.	Select sample, obtain measurement of quality characteristic, compute $\bar{X}$ and S and the lot quality index, C_U for an upper specification limit, or C_L for a lower specification limit.
4. Inspector then:	Compares total number of defectives, C, with acceptance and rejection numbers. If $C \leq a$ the lot is accepted. If $C \geq r$ the lot is rejected.	Compares the lot quality index C (C_U or C_L) with the acceptance constant. If $C_U \geq A_U$ the lot is accepted for an upper specification limit, otherwise rejected. If $C_L \leq A_L$ the lot is accepted for a lower specification limit, otherwise rejected.
5. Severity of Inspection is changed:	By comparing process average with upper or lower limits on the AQL in Table II.	By comparing process quality index C' with severity constants for the AQL in Table II.

(C' is the same as C except that instead of the mean and standard deviation for one lot, the mean and standard deviation for a series of lots is used).

6. Report to BuOrd.

MilStd 105A	OSTD 80
NAVEXOS 2771	NAVORD Form 1707A, 1707B

This then is the procedure for applying OSTD 80.

The use of Forms 1707A and 1707B greatly eases the burden of the inspector in making his decision and reporting his decision by providing for a step by step recording of all the above items. Examples of these forms filled out are given. The data is imaginary, of course.

As already stated, for a double specification limit, the above procedure would be applied twice, once for each limit. In addition a MASD criteria would be imposed if the government desires. The standard fully outlines the simple numerical procedure to compute the MASD.

Also included is a procedure for applying a mixed variables attributes sampling plan in those cases where it is known that the product submitted is selected by the manufacturer from a larger quantity of product all of which would not meet specifications, (i.e., where the distribution is known to be truncated), or in such cases as the government may desire. The procedure for the variables test in OSTD 80 is used for acceptance, but should the lot fail, an additional sample to bring the total sample size up to the appropriate attribute requirement is taken and the old data in attribute form plus the new data is used to make a decision to accept or reject based on the corresponding attribute sampling plan in MilStd 105 for tightened inspection.

Standard Deviation and Range

As previously mentioned, the standard deviation is used in OSTD 80 to estimate dispersion of the quality characteristic measurements from the sample observations. One reason for the use of the standard deviation rather than an estimate by means of the average range was to deter indiscriminate use of the Standard, especially when it was first issued. The ease of computing the range as against the comparative complexity in computing the standard deviation by hand would have resulted in less discriminate selection of cases where application of variables inspection was not a decided advantage. Where appropriate variables inspection should be encouraged but there are dangers in indiscriminate use. For example, because of the dependence of variables sampling plans on the assumption of normality, the use of the range would restrict the applicability of the plans to situations where the normality assumption was strongly valid. When the standard deviation is

used, more leniency in this respect may be allowed.

There are other reasons for prefering the standard deviation. While the range is simpler to compute by hand, the standard deviation lends itself more simply to mathematical operations and thus is more adaptable to general analysis, such as the use of the acceptance inspection data for the additional studies previously mentioned. In fact, the use of the range for acceptance inspection raises mathematical difficulties in specifying criteria to change from normal, tightened or reduced inspection, if the criteria is to be based on all the observations over a series of lots as one grand sample, as is done by Wallis in "Selected Techniques (4), by Bowker and Goode in their book (6) and also in OSTD 80.

Computations based on the average range could result in different estimates of the standard deviation depending on the order in which the sample measurements are recorded. It is conceivable therefore that different estimates based on the same set of data could result in a different decision as to the acceptance or rejection of a lot. One order for recording the sample measurements could be decided on, and always retained. The simplest procedure would be to retain the order in which the measurements are taken. Even so, in sampling from a lot as opposed to a continuous production line, no one order is preferable and contention between producer and inspector could ensue where lot decision is at stake. The standard deviation, on the other hand, provides a unique estimate regardless of the order of the sample measurements.

There is no doubt that the range is desirable for many specialized uses, such as control charts, and will have its place in variables acceptance inspection. Perhaps its place is in those cases where there are some advantages of variables inspection over attributes inspection, but the advantages are limited because the cost or importance of the product tested, or the cost of the test procedure, does not heavily outweigh other administrative costs of applying the plan and performing the computations offers difficulties because they must be done by hand. The procedures in OSTD 80 for lot acceptance can still be applied if the standard deviation is estimated from the range, but the protection afforded to the plan will be less stringent than that given by the OC curves in the Standard. The severity criteria would not apply.

Certainly modern day inspection offices should have a computing machine and someone to operate it. In many instances the cost would be a trifling compared to the savings in inspection by variables. Where desk calculators are available it is not clearly demonstrated that computations using the average range to estimate the standard deviation is simpler than that for the orthodox procedure for the standard deviation. It will be interesting to observe what effect the latest desk calculators which provide automatic square root extraction will have on the question of whether the range or standard deviation should be used.

Closing Remarks

I hope that I have been able to adequately summarize for you the main features of inspection by variables in general, and OSTD 80 in particular. Comments and suggestions concerning the standard are always

welcome. Not by way of apology, but as a statement of fact, it is often difficult to anticipate all the possible situations to which applications of ideas will be made. There probably always will be cases arising to which application of OSTD 80 will be desirable but administratively difficult. The standard may be through revisions to meet new needs and to amend old mistakes, but at the same time it may very well be found advisable to change some established procedures to have them conform with the Standard. A common ground can always be found.

The Standard has served the needs of the Bureau of Ordnance, and will continue to serve its purpose pending the preparation of a more comprehensive Military Standard for the needs of the Department of Defense, which has recently been initiated.

Acknowledgement

It is a pleasure to acknowledge the suggestions and comments received from others in the preparation of OSTD 80.

The initial suggestions to prepare a set of variables sampling plans for use by the Bureau of Ordnance was made by Dr. Pabst now Chief Statistician for the Bureau of Ordnance. He also suggested the use of tightened inspection for attributes when mixed variables - attributes sampling plans are used. The idea of using a maximum allowable standard deviation in applying two single limit plans, came from plans prepared for a certain requirement by William Wolman, Statistician, Naval Inspector of Ordnance, Rochester, New York. Helpful comments have been received from Harry Weingarten in the Statistical Methodology Section of the Statistics Branch in the Quality Control Division who also prepared NavOrd Form 1844a and 1844c which appear in the Standard. Catherine Hock and Fred Frishman of the Acceptance and Surveillance Standards Section also provided pertinent suggestions and have made many applications of OSTD 80 in specifications and for other tests and James Parry provided some assistance in computations. Acknowledgement is also due to the Staff at Frankfort Arsenal for comments and to the staff at Naval Ordnance Test Station, Inyokern, California for comments and for the initial idea to use some form of a quality index to express the acceptance criteria. Finally, I wish to thank Dr. John Tukey and Dr. Acheson J. Duncan for their time in reviewing and comment on a draft of the Standard.

Mathematical Appendix

Known and unknown standard deviation plans for single limit test and single sampling

1. For a single limit test for unknown standard deviation the acceptance criteria is

$$\bar{x} + ks \leq U$$

where $\bar{x} = \frac{1}{n}\sum x_i \quad i = 1, 2, \ldots n$

$$s = \left[\frac{1}{n-1}\sum (x_i - \bar{x})^2\right]^{1/2}$$

U is an upper limit beyond which units are considered defective, and k is a determined constant.

The constant k and the sample size n are uniquely determined by the power required of the sampling plan to discriminate between acceptable product, say fraction defective p_1, and unacceptable product, say fraction defective p_2. The power is specified by requiring that product of fraction defective p_1 be accepted with high probability, say, $1-\alpha$ and, product of fraction defective p_2 be accepted with small probability, say β.

2. If x is normally distributed with mean μ and standard deviation σ, then a fraction defective p is defined by

$$p = \int_{k_p}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2} dt$$

where

$$t = \frac{x - \mu}{\sigma} \qquad \text{and} \qquad k_p = \frac{U - \mu}{\sigma}$$

t is a normalized deviate measured in standard deviation units of any value x from the population mean. k_p is a particular value of t where x = U. The integral from k_p to infinity measures the proportion defective p and this relation is tabulated in tables of the normal distribution. It is because this relation between k_p and p is based on the assumption that x, the quality characteristic, is normally distributed that the statement is made that variables sampling plans are based on the assumption of normality. If x were not normally distributed, but its distribution were known and tabulated then this information would be sufficient in determining the acceptance criteria for inspection by variables to give the power of the sampling plan required.

3. The probabilities $1-\alpha$ and β are defined by

$$1-\alpha = \int_{k_{1-\alpha}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} dz_1 \quad \text{and} \quad \beta = \int_{-\infty}^{k_\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_2^2} dz_2$$

Where, representing acceptable product

$$Z_1 = \frac{z - \zeta_1}{\sigma_z} \qquad k_{1-\alpha} = \frac{U - \zeta_1}{\sigma_z}$$

$$\zeta_1 = \mu_1 + k\sigma_1$$

and, representing unacceptable product

$$Z_2 = \frac{z - \zeta_2}{\sigma_z} \qquad k_\beta = \frac{U - \zeta_2}{\sigma_z}$$

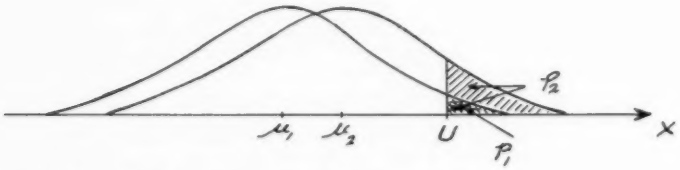
$$\zeta_2 = \mu_2 + k\sigma_2$$

For the sample measurements

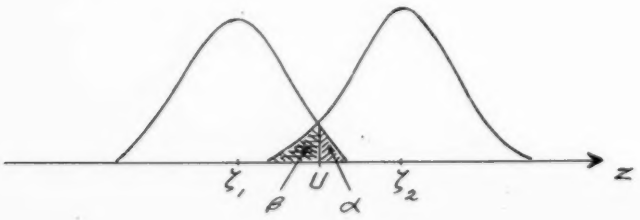
$$z = \bar{x} + ks \quad \text{and} \quad \sigma_z = \sigma\left(\frac{1}{n} + \frac{k^2}{2n}\right)^{1/2} \text{ approx.}$$

Since X is normally distributed, z which is a linear function of X and S will be near normal even though S, for small values of n, is not quite normal. (See paragraph 7). Even if X is not normally distributed z will tend to normality. Thus the relation between $k_{1-\alpha}$ and $1-\alpha$ or k_β and β as shown in tables of the normal distribution will still be valid. It is for this reason that statements are made that consistent decisions for lot acceptance can still be made when the quality characteristic is not normally distributed.

4. A pictorial representative of the fraction defective p_1 and p_2 is



The probabilities α and β are illustrated by



5. To determine the constant k and the sample size n we solve simultaneously the equations

$$\frac{U-\zeta_1}{\sigma_z} = k_{1-\alpha}$$

and

$$\frac{U-\zeta_2}{\sigma_z} = k_\beta$$

since both conditions must hold. In terms of the definitions for ζ and σ_z these equations can be written as

$$U-\mu_1-k\sigma_1 = k_{1-\alpha}\sigma_1\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

$$U-\mu_2-k\sigma_2 = k_\beta\sigma_2\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

Multiplying through by $\frac{1}{\sigma_1}$ or $\frac{1}{\sigma_2}$ as appropriate gives

$$\frac{U-\mu_1}{\sigma_1}-k = k_{1-\alpha}\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

$$\frac{U-\mu_2}{\sigma_2}-k = k_\beta\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

and finally

$$k_{p_1}-k = k_{1-\alpha}\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

$$k_{p_2}-k = k_\beta\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

The ratio of the two equations is

$$\frac{k_{p_1}-k}{k_{p_2}-k} = \frac{k_{1-\alpha}}{k_\beta}$$

Their difference is

$$k_{p_1}-k_{p_2} = -(k_\alpha+k_\beta)\left(\frac{1}{n}+\frac{k^2}{2n}\right)^{1/2}$$

where $-k_\alpha$ is written for $k_{1-\alpha}$

From the ratio of the two equations the solution for k is

$$k = \frac{k_\alpha k_{p_2}+k_\beta k_{p_1}}{k_\alpha+k_\beta}$$

and from their difference $n = \left(1+\frac{k^2}{2}\right)\left(\frac{k_\alpha+k_\beta}{k_{p_1}-k_{p_2}}\right)^2$

This is the solution for k and n where the standard deviation of x is not known.

Where the standard deviation is known

$$z = \bar{x} + k\sigma$$

and

$$\sigma_z = \frac{\sigma}{n^{1/2}}$$

This does not affect the solution for k which remains the same but n' for known standard deviation is

$$n' = \left(\frac{k_\alpha + k_\beta}{k_{p_1} - k_{p_2}} \right)^2$$

Thus the percent saving in sample size where the standard deviation is known is

$$\left(1 - \frac{n'}{n}\right) \times 100\%$$

6. The operating characteristic curve for unknown standard deviation plans is computed from the equation

$$k_p - k = k_{L_p} \left(\frac{1}{n} + \frac{k^2}{2n} \right)^{1/2}$$

where L_p is the probability of acceptance for fraction defective p.

For known standard deviation plans

$$k_p - k = k_{L_p} \left(\frac{1}{n} \right)^{1/2}$$

is used.

7. The above derivation is based entirely on normal distribution theory which is generally adequate for n greater than 10. Actually z has a non-central t distribution which approaches normality as n increases. For an exact treatment based on the non-central t distribution see references (3) and (4).

References

(1) Military Standard 105A, Sampling Procedures and Tables for Inspection by Attributes.

(2) Jennett, W.J., and Welch, B.L.: "The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale," Supplement to the Journal of the Royal Statistical Society, Vol. 6 (1939), 80-88.

(3) Johnson, N.L., and Welch, B.L.: "Applications of the Non-central t-Distribution," Biometrica, Vol. 31 (1940), 362-389.

(4) "Selected Techniques of Statistical Analysis, Eisenhart, Hastay and Wallis, McGraw-Hill Book Company, Inc., New York, 1947, Chapter I.

(5) "Acceptance Sampling--A Symposium," The American Statistical Association, Washington, D.C., 1950, 117-122.

(6) Bowker, A.H., and Goode, H.: "Sampling Inspection by Variables," McGraw-Hill Book Company, Inc., New York, 1952.

(7) "Sampling Inspections," Freeman, Friedman, Mosteller and Wallis, McGraw-Hill Book Company, Inc., New York, 1948.

Table A

Comparison of Sample Size Required in Variables and Attributes Sampling Plans to Obtain Similar Operating Characteristics

	ATTRIBUTES Type of Sampling Plan			VARIABLES Type of Sampling Plan			
Sample Size Letter of Sampling Plan (1)	Single	Multiple Ave. sample size (ASN) at AQL (2)		Unknown Std. Dev. Single	Double Ave. sample size (ASN) at AQL (2)		Known Standard Deviation
		Min	Max		Min	Max	
B	10	6	10	7	5	7	5
C	15	9	13	10	8	9	6
D	20	14	17	13	11	11	7
E	30	17	26	16	13	13	9
F	40	24	35	20	16	17	11
G	55	29	47	25	20	20	13
H	75	38	90	35	27	28	16
I	115	54	131	50	36	38	20
J	150	73	174	60	47	47	24
K	225	96	230	70	51	53	28
L	300	128	306	85	65	67	32
M	450	223	518	100	76	76	36
N	750	348	631	125	92	94	40
O	1500	783	1324	200	151	154	45

(1) Attribute sampling plans are from "Sampling Inspection" By Columbia Research Group.

(2) ASN values are largest and smallest for a series of sampling plans with different AQL's.

Source: "Sampling Inspection by Variables", Bowker & Goode, Tables 3.1, 3.2, 6.1

TABLE B.--Comparison of MilStd 105A and OSTD 80 for Sample Size and *LTPD

LOT SIZE	LEVEL II SAMPLE SIZE		AQL .10		AQL .25		AQL 1.0		AQL 2.5		AQL 4.0		AQL 6.5	
	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80	MIL STD 105A	OSTD 80
2-8	2													
9-15	3												57	
16-25	5										38		↑	
26-40	7	5						< 28	27	< 34.7	↑	< 39	↓	< 44.3
41-65	10								↑		↓		35	
66-110	15	10		6.85		9.31	12.6	15.2	↓	21.2	25	25.4	32	30.9
111-180	25						↑		14.2		20		25	
181-300	35	20		3.10		4.69	↓	9.06	12.9	14.1	18.5	18.0	25	22.8
301-500	50				4.5		7.6		13		16		22	
501-800	75				↑		6.9		10.2		13		19	
801-1300	110	35	2.1	1.82	↓	2.98	6.1	6.42	9.3	10.8	12	14.1	16.2	18.7
1301-3200	150		↑		2.6		5.6		8.8		11.6		15.8	
3201-8000	225	70	↓	1.03	2.5	1.82	4.0	4.44	7.5	8.02	10.1	10.9	14	15.1
8001-22000	300		1.25		2.0		3.8		6.8		9.1		13	
22001-110000	450	130	1.20	.70	1.7	1.31	3.5	3.46	6.1	6.62	8.3	9.3	11.9	13.1
110001-550000	750		.88		1.4		2.8		5.3		7.2		10.6	
550001-over	1500	200	.65	.56	1.0	1.09	2.2	3.02	4.3	5.94	6.3	8.4	9.4	12.1

*LTPD is the percent defective for which the probability of acceptance is .10.
The values are given in the body of the table. Those for MilStd 105A were read from the OC curves in that Standard.

TABLE I.--Summary of Characteristics of the Variables Sampling Plans

Sample size code letter	Sample size	Risk of Rejecting AQL Product	Acceptable Quality Level (AQL)					
			.10	.25	1.0	2.5	4.0	6.5
			Percent defective for which the risk of acceptance is .10					
A----------	5	.10	*	*	28.1	34.7	39.0	44.3
B----------	10	.10	*	9.31	15.2	21.2	25.4	30.9
C----------	20	.08	3.10	4.69	9.06	14.1	18.0	22.8
D----------	35	.06	1.82	2.98	6.42	10.8	14.1	18.7
E----------	70	.04	1.03	1.82	4.44	8.02	10.9	15.1
F----------	130	.02	.70	1.31	3.46	6.62	9.26	13.1
G----------	200	.01	.56	1.09	3.02	5.94	8.43	12.1

* No sampling plan provided

TABLE III.--Lot Size and Sample Size Code Letters for Three Levels of Normal or Tightened Inspection and for Reduced Inspection

Lot size	Sample Size Code Letter			
	Normal or tightened inspection			Reduced inspection
	Inspection Level			
	I	II	III	
Less than 41----------	A	A	B	A
41 - 110----------	A	B	C	A
111 - 300----------	B	C	D	B
301 - 1300----------	C	D	E	B
1301 - 8000----------	D	E	F	C
8001 - 110,000----------	E	F	G	D
110,001 and up----------	F	G	G	E

TABLE II (a).--Severity Constants For Change In Severity Of Inspection For Upper Specification Limits U

Number of Measurements Included In Estimated Quality Index	Acceptable Quality Levels--AQL											
	.10		.25		1.0		2.5		4.0		6.5	
	For Change From Normal To Tightened Or Normal To Reduced Inspection											
	T_U	R_U	T_U	R_U	T_U	R_U	T_U	R_U	T_U	R_U	T_U	R_U
100 - 299	12.70	13.59	12.44	13.27	12.01	12.72	11.68	12.31	11.49	12.07	11.27	11.81
300 - 499	12.80	13.43	12.54	13.12	12.10	12.60	11.76	12.20	11.56	11.97	11.34	11.72
500 - 899	12.87	13.34	12.60	13.04	12.15	12.53	11.80	12.14	11.60	11.91	11.38	11.67
900 - 1899	12.93	13.26	12.66	12.97	12.20	12.46	11.85	12.08	11.64	11.86	11.42	11.62
1900 - 3299	12.97	13.22	12.70	12.93	12.23	12.43	11.87	12.05	11.67	11.83	11.44	11.59
3300 - 4699	13.00	13.19	12.72	12.90	12.25	12.41	11.89	12.03	11.69	11.82	11.46	11.57
4700 - 7099	13.02	13.17	12.74	12.88	12.27	12.39	11.91	12.02	11.70	11.80	11.47	11.56
7100 - 9999	13.03	13.16	12.75	12.87	12.28	12.38	11.92	12.01	11.71	11.79	11.48	11.55
10,000 And Up	13.04	13.14	12.77	12.86	12.29	12.36	11.93	12.00	11.72	11.78	11.49	11.54
	For Change From Tightened To Normal Or Reduced To Normal Inspection											
	N_U		N_U		N_U		N_U		N_U		N_U	
100 And Up	13.09		12.81		12.33		11.96		11.75		11.51	

NOTES: Severity Criteria for Upper Specification Limits

For Change from: Normal to Tightened $c'_U \leq T_U$ Normal to Reduced $c'_U \leq R_U$

Tightened to Normal $c'_U \geq N_U$ Reduced to Normal $c'_U < N_U$

N_U, R_U, T_U, : Severity Constants, see paragraph 9 and NAVORD Form 1707b (Rev.)

c'_U : Process Quality Index. see paragraphs 8.3 and 8.4.3.

TABLE IV.--Master Table for Normal, Tightened, or Reduced Inspection
Sample size and acceptance constants for single lot decision

Sample size code letter	Sample size	For upper specification limits U Acceptable quality levels (AQL) for normal or reduced inspection					
		.10	.25	1.0	2.5	4.0	6.5
		A_U	A_U	A_U	A_U	A_U	A_U
A	5	↓	↓	11.45	11.18	11.02	10.83
B	10	↓	12.07	11.68	11.38	11.21	11.01
C	20	12.45	12.22	11.81	11.50	11.31	11.11
D	35	12.54	12.30	11.88	11.56	11.38	11.17
E	70	12.64	12.40	11.97	11.64	11.45	11.24
F	130	12.70	12.45	12.01	11.68	11.49	11.27
G	200	12.73	12.48	12.04	11.70	11.29	11.51
		.25	1.0	2.5	4.0	6.5	xxx
		Acceptable quality levels (AQL) for tightened inspection					

Sample size code letter	Sample size	For lower specification limits L Acceptable quality levels (AQL) for normal or reduced inspection					
		.10	.25	1.0	2.5	4.0	6.5
		A_L	A_L	A_L	A_L	A_L	A_L
A	5	↓	↓	8.55	8.82	8.98	9.17
B	10	↓	7.93	8.32	8.62	8.79	8.99
C	20	7.55	7.78	8.19	8.50	8.69	8.89
D	35	7.46	7.70	8.12	8.44	8.62	8.83
E	70	7.36	7.60	8.03	8.36	8.55	8.76
F	130	7.30	7.55	7.99	8.32	8.51	8.73
G	200	7.27	7.52	7.96	8.30	8.71	8.49
↓ Use first sampling plan below arrow		.25	1.0	2.5	4.0	6.5	xxx
		Acceptable quality level (AQL) for tightened inspection					

NOTES: A_U, A_L - Acceptance Constants
Acceptance Criteria, see paragraph 11 and NAVORD Form 1707a (Rev.)
Tightened Sampling plans are not provided for AQL. 10.

VARIABLES ACCEPTANCE INSPECTION REPORT
SINGLE LOT ACCEPTANCE
NAVORD FORM 1707A (REV. 3-52) (FRONT)

REPORT SYMBOL ORD-209
SHEET _____ OF _____ SHEETS

INSPECTION OFFICE	TESTING ACTIVITY	DATE OF TEST
CONTRACTOR	CONTRACT NUMBER	LOT NUMBER
UNIT OF PRODUCT INSPECTED (Description)		MARK / MOD
SPECIFICATION NUMBER AND REVISION	ORDNANCE CLASSIFICATION OF DEFECTS, NO. AND REV	END PRODUCT OF CONTRACT (If Specified)

SEVERITY OF INSPECTION (Check one)
[X] NORMAL [ ] REDUCED [ ] TIGHTENED

LOT SIZE: 100

TEST IDENTIFICATION (for Reference to Requirement in Specification, by paragraph, or QCD)

LINE	VALUES	QUALITY CHARACTERISTICS FOR TEST				
		ACCEPTABLE QUALITY LEVELS				
		U 1.0	U	U	U	U
		L 1.0	L	L	L	L
1	SAMPLE SIZE N	10				
2	SUM OF MEASUREMENTS Σx	103.31				
3	SUM OF SQUARED MEASUREMENTS Σx^2	1070.11				
4	CORRECTION FACTOR $CF = (\Sigma x)^2 \div N$	1067.30				
5	SUM OF SQUARES $SS = \Sigma x^2 - CF$	2.81				
6	VARIANCE $V = SS \div (N - 1)$	.312				
7	LOT STANDARD DEVIATION $S = \sqrt{V}$	.559				
8	LOT MEAN $\bar{X} = \Sigma x \div N$	10.39				
9	UPPER SPECIFICATION LIMIT U	12.50				
10	LOWER SPECIFICATION LIMIT L	7.00				

ACCEPTANCE CRITERIA FOR EACH QUALITY CHARACTERISTIC

(a) ACCEPTANCE CRITERIA UPPER SPECIFICATION LIMIT $C_U \geq A_U$

11	ACCEPTANCE CONSTANT A_U	11.68				
12	LOT QUALITY INDEX $C_U = \frac{U - \bar{X}}{S} \div 10$	13.88				

(b) ACCEPTANCE CRITERIA LOWER SPECIFICATION LIMIT $C_L \leq A_L$

13	ACCEPTANCE CONSTANT A_L	8.32				
14	LOT QUALITY INDEX $C_L = \frac{L - \bar{X}}{S} \div 10$	6.06				

(c) ACCEPTANCE CRITERIA FOR STANDARD DEVIATION $S <$ MASD
(Applicable only if the maximum allowable standard deviation, MASD, is specified by the Government. Write NONE if not applicable.)

15	MASD	1.38*				

ACTION TAKEN	INSPECTOR'S SIGNATURE	DATE
Lot Accepted		

* Computed in accordance with directions found in OSTD 80.

VARIABLES ACCEPTANCE INSPECTION REPORT
FOR SEVERITY OF INSPECTION
NAVORD FORM-1707b *(Rev. 3-52)*

REPORT SYMBOL-ORD-209

INSPECTION OFFICE	PERIOD INCLUDED FROM _____ TO _____
CONTRACTOR / CONTRACT NUMBER	LOT NUMBERS *(First and Last inclusive)*
UNIT OF PRODUCT INSPECTED *(Description)*	MARK / MOD
SPECIFICATION NUMBER AND REVISION	ORDNANCE CLASSIFICATION OF DEFECTS, NO. AND REV.
END PRODUCT OF CONTRACT *(If applicable)*	

SEVERITY OF PAST INSPECTION *(Check one)*
[X] NORMAL [ ] REDUCED [ ] TIGHTENED

NUMBER OF LOTS: 20

LINE	VALUES	QUALITY CHARACTERISTICS FOR TEST			
		ACCEPTABLE QUALITY LEVELS			
		U 1.0	U	U	U
		L 1.0	L	L	L
1	SUM OF LINE (1) NAVORD FORM 1707a: ΣN	200			
2	SUM OF LINE (2) NAVORD FORM 1707a: $\Sigma\Sigma X$	2181			
3	SUM OF LINE (3) NAVORD FORM 1707a: $\Sigma\Sigma X^2$	23881			
4	CORRECTION FACTOR $CF' = (\Sigma\Sigma X)^2 \div \Sigma N$	23784			
5	SUM OF SQUARES $SS' = \Sigma\Sigma X^2 - CF'$	97			
6	VARIANCE $v' = SS' \div (\Sigma N - 1)$	.487			
7	PROCESS STANDARD DEVIATION $S' = \sqrt{v'}$	.698			
8	PROCESS MEAN $\bar{x} = \Sigma\Sigma X \div \Sigma N$	10.50			
9	UPPER SPECIFICATION LIMIT U	12.50			
10	LOWER SPECIFICATION LIMIT L	7.00			

SEVERITY CRITERIA FOR EACH QUALITY CHARACTERISTIC

(a) SEVERITY CRITERIA FOR UPPER SPECIFICATION LIMIT

FOR CHANGE FROM: NORMAL TO TIGHTENED $C'_U \leq T_U$; TIGHTENED TO NORMAL $C'_U \geq N_U$; NORMAL TO REDUCED $C'_U \geq R_U$; REDUCED TO NORMAL $C'_U < N_U$

11	SEVERITY CONSTANT	T_U 12.01	R_U 12.72	T_U	R_U	T_U	R_U	T_U	R_U
		N_U 12.33		N_U		N_U		N_U	
12	PROCESS QUALITY INDEX $C'_U = \frac{U - \bar{x}}{S'} \times 10$	12.86							

(b) SEVERITY CRITERIA FOR LOWER SPECIFICATION LIMIT

FOR CHANGE FROM: NORMAL TO TIGHTENED $C'_L \geq T_L$; TIGHTENED TO NORMAL $C'_L \leq N_L$; NORMAL TO REDUCED $C'_L \leq R_L$; REDUCED TO NORMAL $C'_L > N_L$

13	SEVERITY CONSTANT	T_L 7.99	R_L 7.28	T_L	R_L	T_L	R_L	T_L	R_L
		N_L 7.67		N_L		N_L		N_L	
14	PROCESS QUALITY INDEX $C'_L = \frac{\bar{x} - L}{S'} \times 10$	5.02							

ACTION TAKEN *(State severity of inspection to be used for succeeding lots)* — DATE — INSPECTOR'S SIGNATURE

Reduced Inspection

HOW RELIABLE IS YOUR LIFE TEST PROCEDURE?

J. Alfred Davies
General Electric Company

Information Desired From Life Testing

There is a vital need for life testing all types of service products whether they be used as industrial equipment or home appliances. Every manufacturer is faced with the tough problem of satisfying his customerswith the continued satisfactory operation of his products. These problems connected with life of electronic receiving tubes are the bases for this discussion.

In any type of life test we want assurance that the actual life of the product will exceed a specified minimum value. In the case of radio tubes this assurance is expressed by the following indicators:

1. Rate of tube failure during early life;
2. Percent of tube failures during the specified life test period.

Special Sampling Plans Required

The usual acceptance sampling plans cannot be used adequately for the following reasons:

1. Too much time is required for life testing;
2. Tube failures include both attributes and variables;
3. The test is considered destructive;
4. The cost of space and equipment is prohibitive for large samples.

The life test period is 500 to 1000 hours, requiring 3 1/2 to 7 weeks to run. Only a few percent of the tubes are expected to fail during the test period. Further, no accelerated test has been developed to date. In connection with reducing the time of life test, Warren R. Purcell[1] of The Brown Company has developed a clever life test procedure for incandescent lamps by using the median life of the samples on an accelerated life test requiring only a few hours.

In line with the classification of tube failures, Mr. Wally P. Koechel[2] of Tung-Sol is using a reject limit for the average of the major electrical characteristic as a means for acceptance or rejection. He uses previous life test history to set up a declining process level and corresponding reject limits at various life test hours. However, each electrical characteristic and attribute failure must be considered separately.

Our problem is not physical, but statistical. It is to give adequate assurance of specified tube life while minimizing the total life test socket hours.

Tube Failures and Failure Curves

A composite of the findings in the radio tube industry shows that there are two major causes of tube failure during life. A tube may become a failure either abruptly or gradually. The abrupt type of

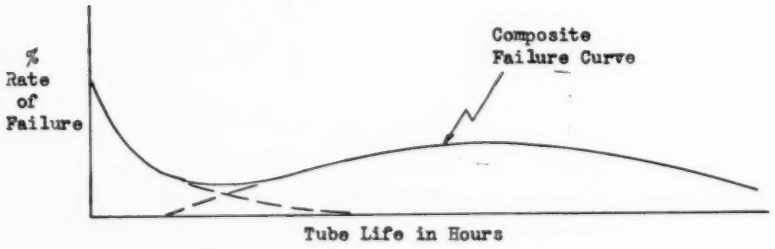
failure is an "inoperative" which has several other names, such as sudden death or catastrophic failure, etc. It is caused by an open weld or burn out of the heater, an electrical short, or a leak in the envelope resulting in loss of vacuum. The other type of failure results from the gradual decline of tube activity such as wear-out of cathode emission.

Table 1 summarizes the main features of these defectives. It should be noted that the attribute defects occur early in life and follow an exponential rate of failure curve, whereas the variable type of defects occur later and have an extremely skewed distribution. consequently the composite rate of failure curve is somewhat as shown in Figure 1.

Table 1. Life Test Defectives, Radio Tubes

Classification		Nature of Defective	Usually Occur	Rate of Failure Curve
Statistical	Engineering			
1. Attributes	Mechanical, Inoperative	Open circuit, electrical short, loss of vacuum	Early in life	Exponential
2. Variables	Electrical, Wear-out	Failure to meet any one of the specified L T endpoints	Later in Life	Extremely skewed distribution

Figure 1. Typical Rate of Failure Curve for Receiving Tubes



This curve shows a period of infant mortality, a level-off period and then a gradual increase until finally dropping off. (It is of interest to note that this mortality curve of tubes is very similar to that for human beings - - even to the normal and the accidental deaths.) The scales have been deliberately left out in order to illustrate merely the general nature of the curve for receiving tubes.

Exactly the same type of mortality curve was found for television tubes by Mr. Richard Wilde[3] of Sylvania. Also, Mr. A. J. Heitner and Mr. A. L. Wilson[4], Sylvania, reported similar results of their investigation of the life characteristics of subminiature receiving tubes.

Types Of Life Test Sampling Methods

It is necessary to consider both types of tube failure in any method of life test evaluation. Three possible approaches to this evaluation as recommended by Mr. Wilde are;

1. Treat each of the types of failures separately in relation to its expected failure curve;
2. Combine both types of failures and use one density function for evaluating them;
3. Use some non-parametric method for evaluating the combined failures.

An ingenius parametric method has been devised by Mr. Lee Harris[5] of General Electric Company. Essentially it is an application of the point binomial. It is presented in tabular form showing the sample size required for life test to be sure, with probability P%, that fewer than k% of future units will fail in a time shorter than the shortest life in the sample. However, for values of k up to 1% failures, the sample size is several hundreds or thousands at the 95% assurance level. So we conclude that the nonparametric methods are not economic and will discuss only those life test procedures based on the first two approaches above. The methods currently being used or considered in the radio tube industry are;

1. The MIL-E-1B rating method;
2. The JTC-11, Committee on Sampling Procedure, method for reliable tubes;
3. The General Electric method for 5-Star tubes;
4. The Military Specification method for military control tubes.

The purpose of these methods is to accept or reject production lots of tubes on the basis of percent failures from 0-500 hours of life. None of these procedures estimates mean life.

Life Test Conditions

The actual procedures for life testing are similar for each one of the sampling methods, so the conditions for life testing receiving tubes are given here merely as background information concerning the physical operations.

Life testing is conducted by intermittent operation under specified voltage conditions, having 12 to 25 cycles of on-and-off periods per 24 hour period. The specified total operating time is 500 hours per sample.

MIL-E-1B Life Test

The lot size is usually one week's production, but is left to the option of the manufacturer. The sample size is as specified for each tube type by groups as follows;

Group A - not less than 5 nor more than 10 tubes per week;
Group B - not less than 3 nor more than 5 tubes per week;
Group C - not less than 1 nor more than 3 tubes per week;
Group D - not less than 1 nor more than 3 tubes per month.

A tube is declared a failure when it has first reached its Life Test End Point or the tube suddenly becomes inoperative. Practically all of

the receiving tube types which are supplied to the Services are rated in group A, so the sample size for conventional tube types is from 5 to 10 tubes per week's production. The sample is rated in per cent of total specified operating time during which the tubes remain operative. The minimum rating of any sample for acceptance of the lot is 80%.

The life test rating equation for this method is

(1) $r\,(\%) = 100\ \Sigma h_i / nT$, where

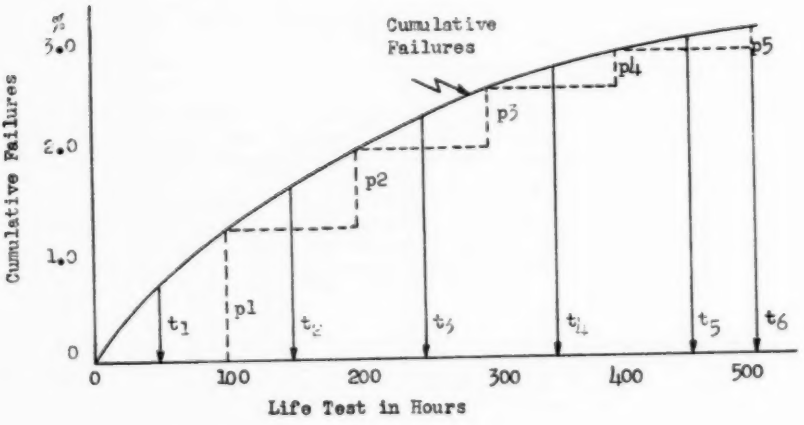
r is the life test rating of the sample in per cent,
h_i is the operative life in hours of the tube until it fails,
n is the number of tubes in the life test sample, and
T is the time in hours specified for life test duration.

To make use of equation (1) we define the reading periods to be those times at which the sample is removed from the life test shelf and is tested. From the integrated rate of failure curve, approximations to the probabilities, p_i, of failure within these intervals are readily obtained by subtraction of the cumulative percentages at successive reading periods. These probabilities, weighted by their respective midpoints, t_i, of each time interval provide an approximation of total tube-hours.

Thus the expected rating can be estimated from the equation, using T = 500,

(2) $r\,(\%) = \Sigma t_i p_i / 5$

Figure 2. Illustration of Life Test Rating



The expected rating for this illustration is 97.98%. Thus,

i	p_i	t_i	$t_i p_i$	
1	.013	50	0.65	
2	.007	150	1.05	
3	.005	250	1.25	$\Sigma t_i p_i$ = 489.90
4	.003	350	1.05	
5	.002	450	0.90	Expected r = 489.90/5 = 97.98%
6	.970	500	485.00	
	1.000		489.90	

The advantages of this MIL-EL-B rating method are: The combination of both electrical and mechanical failures requires only one statistical procedure, and early failures are penalized most severely by the weighting. Disadvantages are: The 80% minimum rating for acceptance is too low for good evaluation, and the failure curve must be known.

JTC-11 Life Test Method

The JTC-11 acceptance sampling plan for life test has been designed to treat each type of tube failures separately. Essentially this procedure consists of three separate, but integrated, tests: 1. Check of early life failure rate; 2. Control of later life sporadic failures; 3. Control of electric items. There is a relatively large sample A operated for the first 100 hours and a sample B consisting of 20 tubes to be continued from 100 to 500 hours. The purpose of sample A is to evaluate the early life failures which are mainly attributes. Sample B performs a test by variables for the per cent decline of the major electrical characteristic (s). It also serves as a rough check on attributes failures between 100 and 500 hours.

The sampling plans for sample A are obtained from the MIL-STD-105A tables and provide for Normal, Reduced, and Tightened sampling, using either Single or Double sampling plans. As an example, for an AQL of 0.65% and lot size 8001 or more, Normal sampling requires the sampling plan for Life Test Sample A to be either n = 300, c = 5 for Single sampling, or n_1 = 200, n_2 = 400, c_1 = 3, and c_2 = 7 for Double sampling The sampling plan using sample B for attributes is n = 20 and c = 2 for all lot sizes. For the test by variables the average value of each life test characteristic for Life Test Sample B at 500 hours may not differ from the average value at 0 hours by more than the allowable shift specified.

The operating characteristic curves for the two attributes portions of this life test are shown in Figure 3. The OC curve for the combination is essentially the same as the 0-100 hours' curve. It derives from the productive probabilities of the other two separate OC curves, assuming that the slight dependence between the two attributes curves is negligible for all practical purposes.

Advantages of this method are that it tests the two types of tube failures by using a large sample to evaluate by attributes the early-life, mechanical defects and a small sample to test the later-life, electrical defects by variables. This method is also quite general and can be used to evaluate any tube type regardless of its failure rate or specified AQL. Its disadvantages are the lack of sensitivity

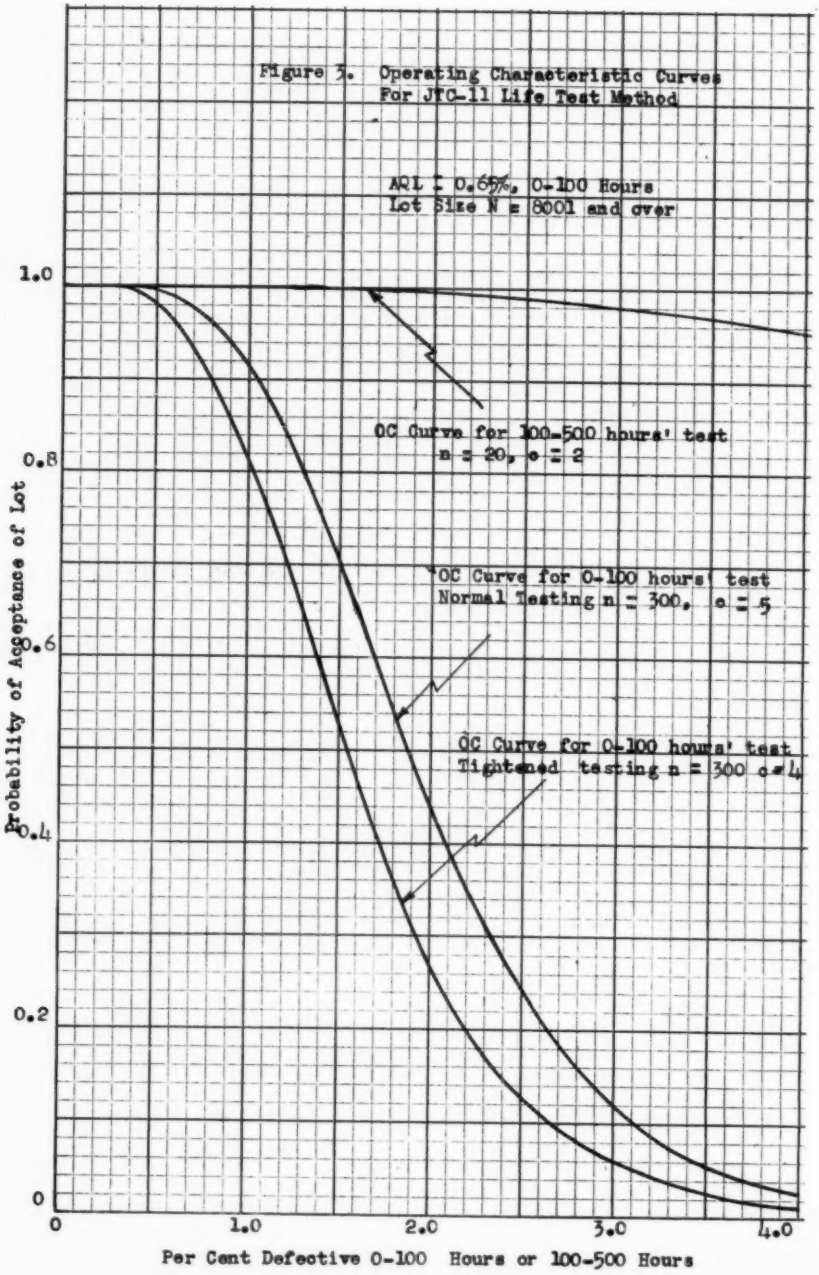


Figure 3. Operating Characteristic Curves For JTC-11 Life Test Method
AQL = 0.65%, 0-100 Hours
Lot Size N = 8001 and over
OC Curve for 100-500 hours' test n = 20, c = 2
OC Curve for 0-100 hours' test Normal Testing n = 300, c = 5
OC Curve for 0-100 hours' test Tightened testing n = 300 c = 4
Probability of Acceptance of Lot
1.0
0.8
0.6
0.4
0.2
0
0
1.0
2.0
3.0
4.0
Per Cent Defective 0-100 Hours or 100-500 Hours

of the attributes test for sample B and the large number of tubes required to be tested to 100 hours. The latter objection is made on the basis of equipment and space required rather than on the amount of testing effort.

General Electric Life Test Method

This method is a combination of the MIL rating method (using a higher per cent minimum requirement) and the JTC-11 method for variables. The same rating procedure is used as described previously. However, the sample size is 20 tubes for the entire 500 hours, provisions are made for Normal or Stricter sampling, depending on the process average, and the product rating is specified at 98.5% or better. A minimum acceptable rating of 95% is permitted for any sample, providing the process average of the last 10 samples is 97.5% or better. Otherwise, the minimum permits only one tube failure in the sample between 0 and 100 hours, whereas the 96% rating does not permit any failure during this early period.

On the basis of a constant failure rate of 0.6% per hundred hours this plan has been devised to evaluate a product which averages 98.5%, or better, operative life. At 0.6%/100 hours the total failures to 500 hours is 3.0%. This will be weighted by 1/2, the average life for the interval. Thus, the tube - hours lost is 1.5%, giving an expected rating of 98.5%. This assumption of a constant failure rate gives the relation between rating and per cent failures:

(3) $F'(\%) = 100 - p/2$, where

p is the cumulative per cent failures 0 to 500 hours.

Empirical studies of drawings from a sampling bowl show that this estimation equation holds very well for a constant failure rate, and also that the standard error of the rating is approximately inversely proportional to the square root of the sample size.

Table 2 summarizes the bowl drawings and calculations of average rating, standard deviation, and percent of lots accepted for various combinations of sample sizes and failure rates. If the early failure curve still remains, equation (3) no longer applies. The early failures are weighted more heavily, and since there are proportionately more of them, much higher rejection rates occur. Consequently, the operating characteristic curve for the assumed condition of constant failure rate appears more lenient than it actually is. Strictly speaking, it is the upper limiting bound of OC curves that could be constructed by assuming different types of concave failure rate curves. The operating characteristic curves for Normal and Stricter sampling, assuming a constant rate, are shown in Figure 4, and the points marked with "x" and "χ" show the corresponding acceptance figures from a population having an exponential failure rate.

This life test rating plan when used in conjunction with some control procedure for minimizing early life failures gives the same advantages as the MIL-E-1B and the JTC-11 methods. The relatively small sample required for a life test having good efficiency represents further advantages of economy and practicability. Its disadvantages are: No particular one OC curve can be used without specifying the

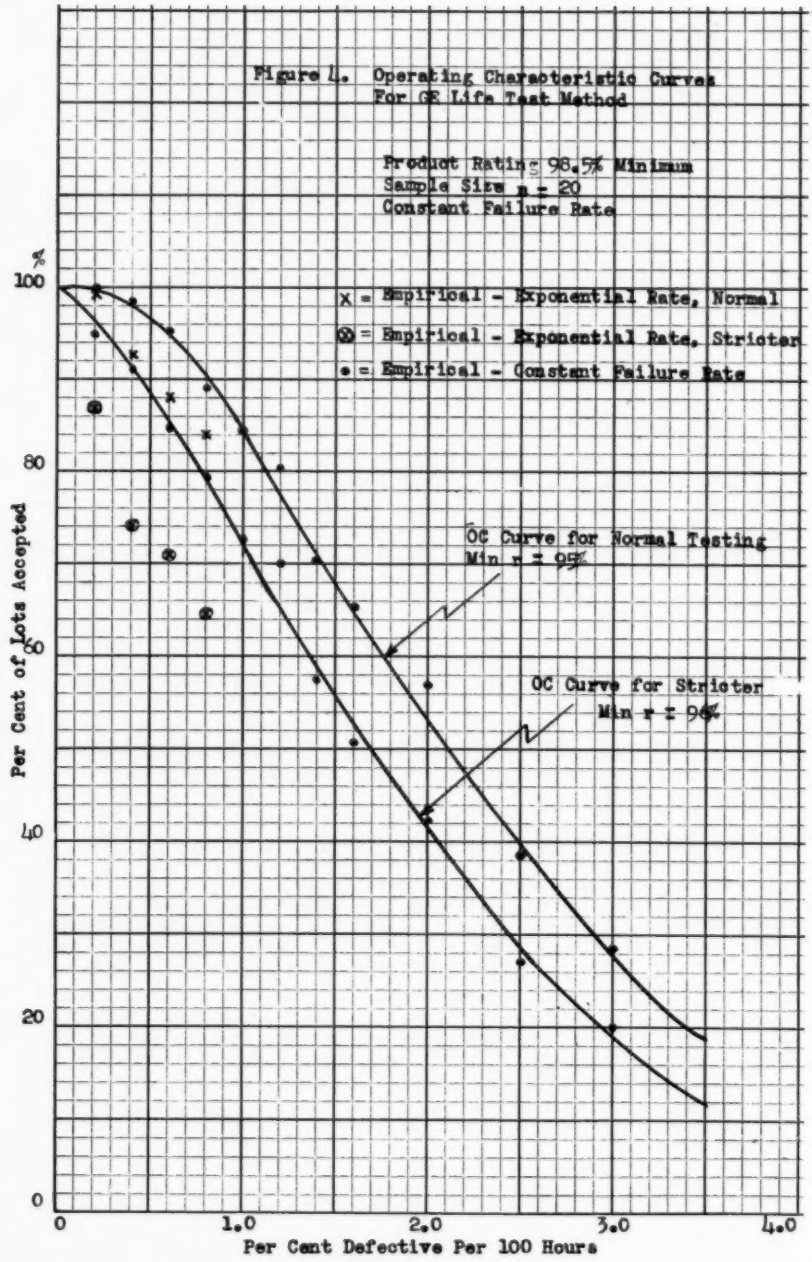


Figure 4. Operating Characteristic Curves For GE Life Test Method
Product Rating 98.5% Minimum
Sample Size n = 20
Constant Failure Rate
X = Empirical - Exponential Rate, Normal
⊗ = Empirical - Exponential Rate, Stricter
• = Empirical - Constant Failure Rate
OC Curve for Normal Testing Min r = 95%
OC Curve for Stricter Min r = 96%
%
100
80
60
40
20
0
Per Cent of Lots Accepted
0
1.0
2.0
3.0
4.0
Per Cent Defective Per 100 Hours

rate of failure curve, and for Stricter sampling it does not provide for a random failure up to 100 hours.

Military Specification Life Test

This is a recently proposed specification which includes the following separate life tests; 1. Stability of electrical characteristic for the first hour; 2. Survival rate during the first 100 hours; 3. Control at 500 hours, inoperatives and several electrical items separately by attributes; Similar control at 1000 hours.

The Stability test is performed on a sample of 50 tubes to a 1% AQL. The Survival rate test is made in accordance with MIL-STD-105A sampling procedure at 0.65% AQL, inspection level II. This is identical to the JTC-11 method except the AQL is specified. The OC curve is the same as Figure 3, 0-100 hours.

However, the intermittent 500-hour test is different. It permits either single or double sampling with acceptance numbers specified for each separate test item. A single sample of 20 tubes is required and the acceptance number is either 1 or 2 defectives per test item. Double sampling uses the same first sample of 20 with acceptance number 1 or 2 per test item and a second sample of 40 with acceptance number 3 or 5, respectively. Acceptance numbers of 4 and 8 are used for all types of defectives in the first or combined samples.

The OC curves for the 500-hour test are shown in Figure 5. These curves have been plotted to the abscissa scale in percent defective per 100 hours.

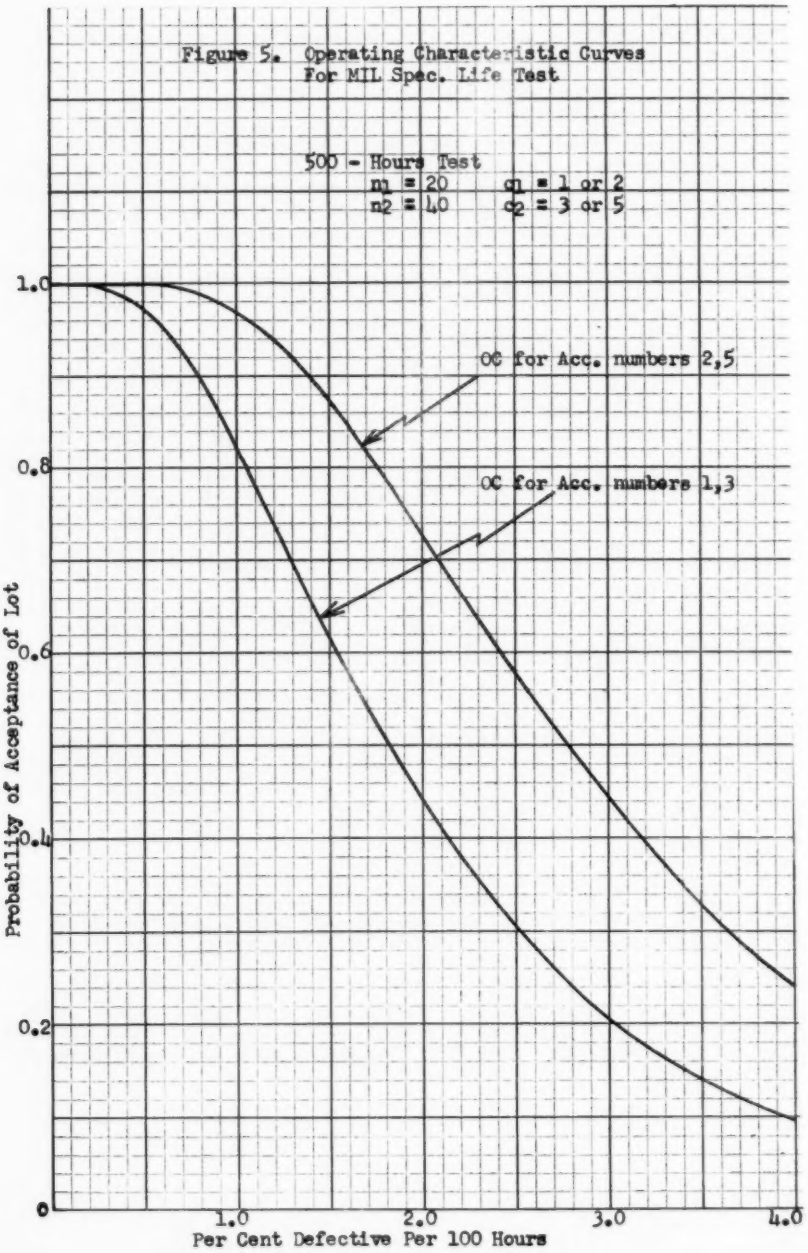
Advantages of this plan are the same as the JTC-11 method, except for flexibility of the 100- hour test, and the 500-hour test gives better control. Disadvantages are; the composite OC curve cannot be established unless the relative proportion of each type of defective is known, and the present classification of defects makes each electrical test as important as the inoperatives.

Conclusion

Figure 6 gives a comparison of the OC curves for the methods of JTC-11, Military Specification, and GE. All curves have been drawn on a rate/100 hours scale to make the best comparison.

From an overall view of quality assurance and cost of quality, the rating method can be well recommended. We believe that the rating procedure modified by a tighter control of later-life failures is worth further investigation, since its theoretical possibilities have hardly been touched.

Figure 5. Operating Characteristic Curves For MIL Spec. Life Test



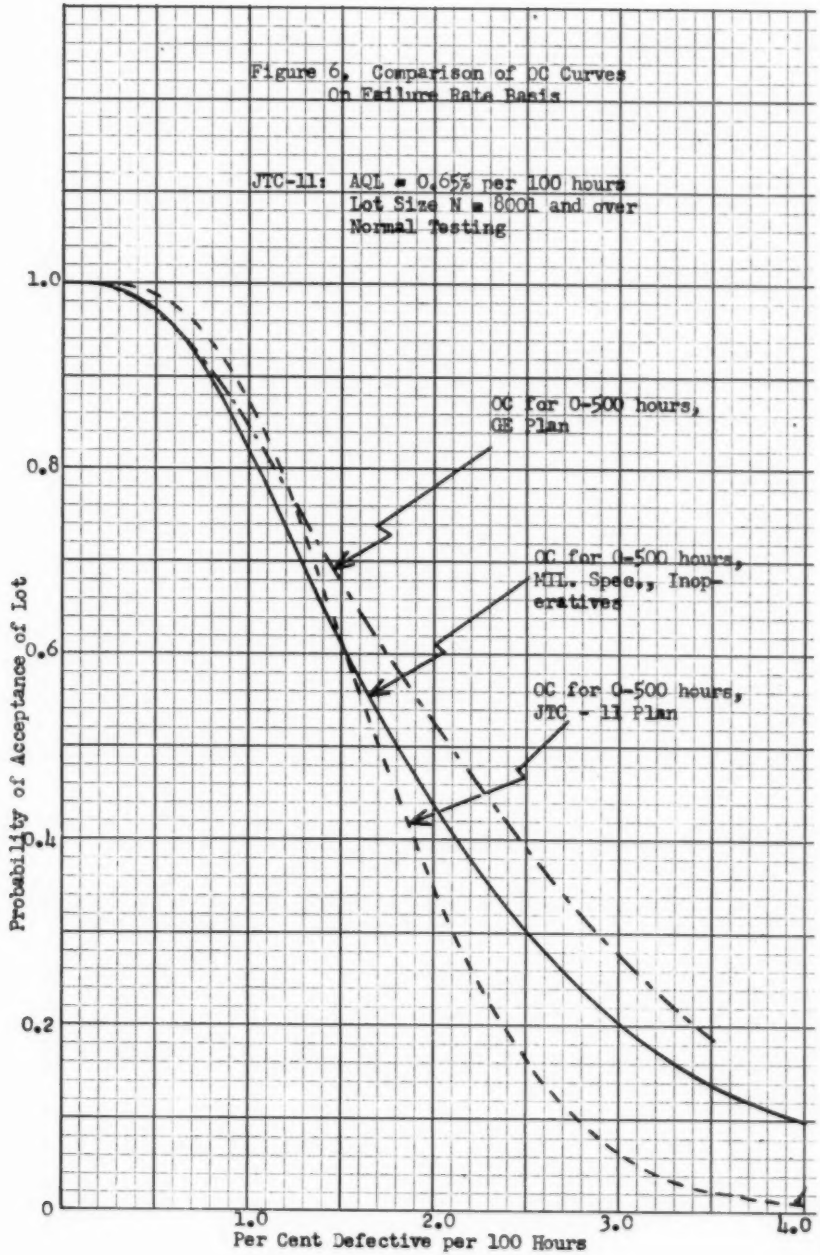


Figure 6. Comparison of OC Curves
On Failure Rate Basis
JTC-11: AQL = 0.65% per 100 hours
Lot Size N = 8001 and over
Normal Testing
OC for 0-500 hours,
GE Plan
OC for 0-500 hours,
MIL. Spec., Inoperatives
OC for 0-500 hours,
JTC - 11 Plan
Probability of Acceptance of Lot
1.0
0.8
0.6
0.4
0.2
0
1.0
2.0
3.0
4.0
Per Cent Defective per 100 Hours

Table 2. Empirical Life Test Ratings
From Bowl Drawings

Failure Curve		Sample	Number of	Empirical		Lots Accepted	
Nature	Rate Per 100 Hours	Size	Samples	$\bar{r}$	σ_r	Nor.	Str.
	%			%	%	%	%
Constant	0.6	5	500	99.44	4.42	-	-
Constant	0.6	10	250	99.45	2.96	-	-
Constant	0.6	20	250	99.48	2.15	95.6	85.6
Exponential	0.2 Avg.	20	250	99.19	2.06	99.2	86.8
Exponential	0.4 Avg.	20	250	98.32	2.61	91.6	74.0
Exponential	0.6 Avg.	20	250	97.59	2.98	88.0	70.4
Exponential	0.8 Avg.	20	250	97.33	3.06	84.4	64.4
Constant	0.2	20	250	99.39	1.31	99.6	94.4
Constant	0.4	20	250	98.97	1.72	98.4	90.8
Constant	0.6	20	250	98.48	2.15	95.6	85.6
Constant	0.8	20	250	98.07	2.45	89.2	79.2
Constant	1.0	20	250	97.39	2.64	84.0	72.4
Constant	1.2	20	250	97.07	3.06	80.4	70.0
Constant	1.4	20	250	96.04	3.39	70.4	57.6
Constant	1.6	20	255	95.74	3.54	65.1	50.6
Constant	2.0	20	250	94.64	4.21	54.8	43.2
Constant	2.5	20	250	93.16	4.38	38.8	27.2
Constant	3.0	20	250	91.73	4.55	25.6	20.0

References

1. Warren R. Purcell, "Saving Time in Testing Life", IQC, March, 1947, pp. 15 - 18

2. W. P. Koechell, unpublished paper, discurred in meeting, 1951, of JTC-11 Committee on Sampling Procedure—Mr. J. R. Steen, Chairman

3. R. D. Wilde, "Nature of Rate-of-Failure Curves," presented at the Summer Statistics Conference, Blue Ridge Assembly, Blue Ridge, N. C. July 2, 1952. Dr. Gertrude Cox, Conference Director.

4. A. J. Heitner and A. L. Wilson, "Analysis of Kew Gardens Premium Tube Life", Sylvania Electric Products Inc., Kew Gardens 15, L. I., May 16, 1951.

5. L. B. Harris, "Evaluation of Life-Test Results," Abstracts from the First Annual Symposium on Statistical Methods, Nov. 7, 1951.

"STATISTICAL METHODS FOR PRODUCT WEIGHT CONTROL"

A.H. Maack - Quality Control Engineer
J.R. Brady - Supervisor of Research and Testing
International Harvester Company
Wisconsin Steel Works

It would probably be well to start out by anticipating some questions which will come to your minds by giving you a brief word picture on the type of steel plant we operate. At several of the meetings held in the midwest many persons were quite surprised to hear that the International Harvester Company operated its own steel mill. Secondly they were surprised to hear that Wisconsin Steel Works is located in the steel producing area of South Chicago. And also that it is the oldest mill in the South Chicago area having been founded in 1875. The plant, however, became a part of the International Harvester Company in 1902.

In order to help visualize our operation it may be well to describe the mill as a completely integrated steel plant. We operate our own coal mines, ore steam ships, three blast furnaces, 148 coke ovens, 11 basic open hearths, a 40" blooming mill, four merchant mills, a cold drawing mill, a heat treating plant, and the necessary service departments to keep such a mill in operation. As to our ingot producing capacity we are rated at 1,000,000 net tons per year or slightly under one percent of the nations capacity. During our last fiscal year we produced 1,042,313 net tons of ingots the greater portion of which was rolled into merchant mill products and totaled 710,000 net tons. Practically all of this product at the present time goes to the parent company or to its direct suppliers.

We maintain a cyclic rolling schedule which enables us to roll most sections approximately every 30 days. Sections for which we have less demand are rolled at approximately 60 or even 90 day intervals. We roll many agricultural shapes some of which are rolled only by us. During a year we produce approximately 3400 different sizes and shapes.

We had been studying the application of Quality Control technique for some time, when in March 1949 our top management requested that we concentrate our effects on our mill products in order to determine possible improvements in processes and product. Our first thoughts were to apply these techniques to bar sizes because this would involve control by measurement and could be done without imposing much additional work on inspection. Also by means of these measurements we could set up a control chart which would follow the process closely and anticipate trends which would tend to produce material outside tolerance, making corrections immediately and therefore keeping the product within specified limits.

Although no particular difficulties had been encountered with our products it was thought that a complete statistical study of our processes might lead to more efficient operation and possibly prove of definite valve in promoting economics for the company as a whole.

The Inspection Department is responsible for all materials leaving the plant meeting the customers ordered requirements. Our Quality Control section was set up independently of Inspection and was made a part of the Research and Testing Department. As such we were a staff organization to offer recommendations to production and reporting through our department head direct to top management. The Inspection Department was however to maintain any charts, take the necessary samples and action. In this way through Inspection we would be able to immediately call to the attention of the production departments serious process deviations. The mill inspector, through the Rolling Supervisor, has direct control of the rolling operation, and as they work as a team, material produced out of tolerance is noted immediately and the necessary adjustments made.

In order to install this Quality Control program for Product Weight Control and to assure getting reliable data, a complete investigation of our sampling set up was necessary. Up to this time our inspection was one of a more or less random nature. By that is meant, the inspector took a certain number of samples which he was required to take for order and heat separation, upset inspection, harden ability etc. The number of checks made for size control was at the inspectors discretion and depended somewhat on the end use of the order; the quality of the steel; and in many cases the number of tons of the size being rolled. The number of these checks depended in many instances on how busy the inspector was watching for other characteristics which came under his jurisdiction. Sampling was usually done from the center of the bar, but samples were taken from other locations if it were suspected that heating was non-uniform and might be causing large variation in section from front to back of bar. Often a true record of what took place during a rolling with respect to dimensional specification was quite lacking.

With the foregoing in mind we started working with inspection on a part time basis gathering data for analysis from a few selected rollings. Some of these rollings were of only a short duration while others lasted for as long as seventeen hours of a single size. The data from these selected rollings was then analyzed and studied to determine the level of operation as well as the amount of variation or spread within it. This was done in order that we could standardize on a sampling plan which would not be overburdening but would still give us an accurate picture of the operation. Were we to require too much the content of the inspectors job would be increased and we would no doubt be required to increase inspection expense.

In this way we varied our sample size and frequency and at the same time checked by 100% inspections many bundles from these same rollings while they were in the warehouse awaiting shipment. Our studies told us that the center of the bars were in most cases representative of the average measurements along the entire length of the bar. Although the front end being slightly hotter when rolled might be somewhat smaller than the back end which has had the maximum chance to cool before being finished to size. This amount of variation depends normally on size, specification, and set up of the mill in which the material is rolled.

Next we had to determine the amount of variation which existed between measurements in our process. This would be caused by a non-uniform temperature between heating furnaces or by momentary delays in the mill during the rolling of the billet. In three of our mills we have two billet heating furnaces which must be kept at the same temperature if the finished size is to be maintained uniform. In these mills it is regular practice to roll two billets from one furnace and then two from the other, to provide a uniform temperature and an even flow of material through the mill. This practice of discharging the furnaces is kept in mind and the orders are charged accordingly as one half of an order goes in one furnace and the other half in the other furnace. This practice can be varied somewhat if heating gets out of line but it is very limited due to the way the material has been charged. All orders, heats, and specifications, must be kept separate. The practice of using two furnaces and discharging two billets must be remembered when setting up a sampling proceedure because both furnaces must be represented in what ever sample is decided upon.

When the billet has been rolled out into a very long bar it is allowed to cool on a large cooling bed before it is cut up into the required lengths. If the bars were not allowed to cool somewhat before shearing we would be unable to maintain straightness and shearing would cause a severe mashing of the bar ends. It is at the time the bars are sheared that our hot shear inspector works enabling us to get samples from any portion of the bar for final cooling and measuring. The samples thus taken are cooled in water before being measured and all our control charts are based on these cold measurements. In the shearing operation we may cut anywhere from one bar at a time on large sizes to twenty at a time on small sizes. From the time the bar leaves the last set of rolls and the time inspection is performed at the shearing operation we may at times have as much as forty or fifty tons of material on the cooling bed.

Some of you will say that this is a post mortem inspection but at high rolling rates with the impossibility of getting samples from fast moving bars, and the fact that customer measurements are all made cold we have accepted this as being as close to the process as it is possible to get. Large variations, if suspected, may be caught however by the

gauger who takes hot measurements from the back ends of the bars. These samples are taken either with a burning torch, hand axe, or are simply made by momentarily holding the bar end up from the table with a tong long enough to obtain measurements. These measurements vary an unknown amount from the center of the bar but do give the gauger and the roller indications as to radical changes in the level of operation. If the gauger suspects from his samples that the level of operation has changed too much he will signal the inspector and a set of fronts, centers, and possibly backs will be taken and cooled in order that adjustments to the mill can be made.

Several months and many thousand measurements later we felt we knew pretty well the normal amount of variation within the process, as well as the level of operation on the many sizes we had studied. During the period we had been collectingour data we had charted $\bar{X}$ and R charts and calculated both process limits and modified or reject limits. At a meeting of management, inspection, and rolling departments we submitted data and made numerous recommendations and it was decided that we should work with the training department and as soon as possible start a training program for inspection, and rolling supervision. This was done within a very short time and we then proceeded with the installation of our program of charting.

It had been agreed at our meeting that we would adopt modified control limits based on the range of our process. Process limits were felt to be impracticable since very few of our ordered tolerances required such concise control. It was also felt that rolling crews working to process limits would, by attempting to hold such close limits lose production, whereas modified limits would insure us a satisfactory product so far as dimensional specification was concerned and would permit maximum production.

The use of modified or reject limits is practical only where the spread of the process (estimated at $\pm 3\sigma'$) is appreciably less than the difference between the two specification limits (X upper tolerance - X lower tolerance). The idea behind the modified control limits being to permit a limited shifting in the process average in cases where the difference between the two specification limits is substantially greater than the spread of the controlled process. It was intended to avoid the cost of stopping production and hunting for trouble when the shifts in process average were not sufficient to cause the production of out of tolerance material. In this way the process can be allowed to vary between the largest acceptable dimension and the samllest acceptable dimension.

From data which we had gathered during our study, we finally agreed with inspection to use a sample size of 4 and a sample frequency of approximately 30 minutes. The samples were to be taken from 4 adjacent bars and would therefore

represent the product of both furnaces. From these data we knew that for the sizes studied the long run average range ($\bar{R}$) was .003 inches.

Using $\sigma' = \frac{\bar{R}}{d_2}$ we can calculate the standard deviation

$$\text{as} = \frac{.003}{2.059}$$

$$= .0014$$

The distance of the $URL_{\bar{X}}$ below the upper specification limit for individual values is

$$3\sigma' - \left(\frac{3\sigma'}{\sqrt{n}}\right)$$

or $$\left[3 - \left(\frac{3}{\sqrt{n}}\right)\right]\sigma'$$

The factor $\left[3 - \left(\frac{3}{\sqrt{n}}\right)\right]$ is called the V for reject limits therefore the

Upper reject limit $_{\bar{X}}$ = X max (upper specification limit) $- V\sigma'$

Lower reject limit $_{\bar{X}}$ = X min (lower specification limit) $- V\sigma'$

for sample size $n = 4$

$$V\sigma' = 1.5 \times .0014$$
$$= .0021$$

Therefore our reject limits or modified control limits as we prefer to call them are set up .002 inches inside the upper and lower tolerance. These hold so long as the range of our process does not get out of control. The control limit for range being calculated as follows:

$$UCL_R = D_4 \bar{R}$$

$$= 2.28 \times .003$$
$$= .00681 \text{ or } .007$$

Chart 1 shows an early study of a rolling of 1-3/16 inch round bar from our number one merchant mill. This chart shows two curves plotted on the average portion of the chart and a single curve in the range portion. It is our regular practice to use this method of plotting on both round and square sections. As our material being rolled in open passes is not perfectly true and therefore on these two sections we chart the average of the largest measurements and the average of the smallest measurements. On the range portion of the chart we show only the largest deviation of the two average curves.

On this chart is shown the modified limits based on

the specification limits for this size as published by the American Iron and Steel Institute. It is these specifications by which we must be governed. These modified limits have been calculated using our long run $\bar{R}$ of .003" and are shown .002" inside specification at ± .008". The regular statistical or process limits when calculated in the usual manner

$$UCL_{\bar{X}} = \bar{\bar{X}} + A_2\bar{R}$$

$$LCL_{\bar{X}} = \bar{\bar{X}} - A_2\bar{R}$$

$$UCL_R = D_4\bar{R}$$

would be placed at +.0054 and +.0012. The limit for range is calculated as .0046.

If these limits were to be imposed on this early rolling they would I believe bring out our reasons for adopting modified control limits. In addition to this we are faced with many short runs of a size and the inspector would be unable to keep up with the mill if he had to change charts and limits for each of these short runs. A little later I will discuss this short run story in more detail.

Another tolerance with which we must concern ourselves is the out of round tolerance. This is not shown as such by this type chart but it is the difference between the high and the low measurement on an individual bar. Although not accurate, a comparison of the two curves on the chart will give a fair indication of the out of round condition.

Studies were made in each of the three mills under Quality control just prior to the time action was taken and charts 2, 3 and 4 show typical rollings at that time. These charts each show about the same picture and although with the amount of off tolerance material they represent it was not of such a serious nature as to cuase trouble with our customers. Each chart shows a level of operation above the mean dimension which was being rolled at that time. Chart number 4 shows a change in control limit due to a change in tolerance for sizes 5/8" in diameter and over.

Chart number 3 representing about 19 hours rolling shows that only 5 out of 38 plottings indicate material on which the smaller dimension of the round is lower than the mean dimension. In this mill it appeared quite evident that we were attempting to use only the upper half of the tolerance, and at the same time the charts show a considerable amount of variation.

Starting in September 1949, and with a greater degree of coverage, charts were installed at the hot shear inspection station in these three mills. Inspection personnel after

their training were assigned the job of keeping the charts current. New charts were put into the mill at the start of the rolling week and a contineous record was kept on all rounds, squares, and flat products, shapes not being included at this time. Checks were still being made in the warehouse to see how our sampling plan was functioning.

When we compared rollings of the same size from month to month we were able to check on our ability to duplicate set up and also our level of operation for the size from month to month. After considerable further study recommendations were made to the effect that the size and therefore the foot weight of our product in many cases was higher than necessary and could be reduced considerably. It became quite apparent that this attempt to roll to the high side of the tolerance was intentional.

It was also recommended that steps be taken at the same time to reduce the large amount of variation which seem very unnecessary.

The attempt to roll to the high side and the excessive variation, is I believe readily understandable when we realize that steel of all classifications, unlike most other products, regardless of specification or how ordered is sold by weight. Also the mill hands get paid on a tonnage basis. Also during the years just prior to this we had rolled large tonnages for cold drawing mills where large stock (to a degree) is more acceptable than stock rolled to the bottom side or under the lower tolerance. Extra stock allowed plenty of clean up in customer plants as well as in subsequent operations in our own. Because of this it was common practice to vary the size of the bar in order to maintain the maximum yield of good steel from the billets provided, and if after obtaining section, croppage was long due to billet variation being on the heavy side the passes would be opened slightly to produce a larger slightly heavier, bar which would reduce the croppage. In this way yields and tonnage rates were kept high and croppage for remelting was low.

Management acted and orders were given to roll to the nominal size regardless of length of croppage. When an attempt was first made to reduce the size of the bars no trouble was encountered with flat products but rounds and squares proved to be more difficult. In some cases it was impossible to roll a smaller round without getting an out of round section. Studies of previous rollings enabled us to tell the level of operation we were maintaining from rolling to rolling. Using this information and the data from the roll shop we were able to recommend changes in allowances for shrinkage as well as sizes to which rolls should be dressed. Effort was made by all concerned to cooperate and at a joint meeting with mill management it was decided to start immediately changing all passes so that bars when cold would vary within tolerance and about the nominal size ordered. This required many changes in the roll shop and necessitated the

purchase of many new plug gauges.

This change was spread over many months and we found that it required a reduction in plug size of as great as .017" in some cases while others required very little change. The most drastic changes were made in the passes maintained for rolling material to be cold drawn. These passes were mostly those for the 1/32 inch sizes. Passes for squares also had to be modified, but passes for flat products and sections required little or no changes.

For a time after each size roll was changed it was apparent that our crops were longer and the cry of low yields and cutting our bonus was heard. This was due to the billets being provided for the top side of the tolerance and therefore overweight for nominal size bars. Adjustments downward in the ordered billet weight by the production control department was required by these changes, but these changes were spread over a considerable period of time and only small losses in overall yield were felt. Now with the majority of adjustments having been made our croppage is no greater than before quality control was installed. Our bars produced under these new methods are considerably closer in weight per foot to those ordered and contain less variation between bars.

Chart number 5 shows a group of sizes similar to those shown in chart number 2 which were rolled after changes were made. The number of samples of each size is not the same on both charts but is representative of the varying amounts of a size rolled from month to month. Although the actual improvement with respect to nominal size only shows a reduction of .0018 inches the material represented is of a much improved quality. The number of points out of control has been reduced from 22 to 2, and the product contains much less variation which we believe is much more acceptable in a competitive market.

Chart number 6 although not showing the same sizes as shown on chart 3 shows what we have been able to do in our number two mill. It is in this mill that we have shown our greatest amount of improvement. There is no question as to the improvement in quality when these two charts are compared. An interesting note on this later chart is that there was a change of turns and a different group of operating personnel setting up the mill for the 1-1/4 inch size. The sizes on number six chart represent the smallest size which we are set up to roll in this mill.

Chart number 7, although not showing the same group of sizes shown on chart 4, shows a control chart representing a more recent performance of this mill. It shows no points out of control and represents a great improvement in practice when compared to the earlier rolling on chart 4. A comparison of these two charts indicates a much closer adherence to nominal size as well as less variation from sample to sample.

It is in this mill that we roll our smallest sizes and improvement was the most difficult.

Chart number 8 represents our first attempt to chart a very close tolerance section. This section has a tolerance of ±.004 of an inch in thickness and is as close, I believe, as any tolerance we maintain on hot rolled bars. The chart shows the control limits set at ±.002 of an inch from nominal size. Although the chart shows 6 out of 39 points out of control, it can be said that we were pleasantly surprised with the results at the end of this rolling. This chart was plotted in the January 1950 rolling of these sections. When we first requested permission to install a control chart on this section, mill management was quite skeptical as to their ability to hold a ±.004", let alone meet a ±.002" as we requested for limits on our control chart for averages. Our training still hadn't convinced them that we were not cutting their tolerance when we put control limits inside what they knew their specification limits were. In order to finally get permission to chart this rolling we installed a chart without limits and promised not to keep after the roller and thus cut production in order to avoid variations and out of control material. Under these conditions we charted the rolling and at its completion returned the chart to the quality control office where modified limits were calculated. These fell as you see them at a ±.002 inches. The mill organization had trouble believing the results of this their first charting on this section.

Chart number 9 shows a later rolling of the above section and although we were surprised at our ability to hold control as well as we did in January 1950 it shows that considerable improvement was still made.

Earlier in this paper I mentioned short runs of a size and the difficulties we would encounter with statistical limits. Even using a modified limit we find the control limits jumping around so badly that it is almost impossible to keep up with them. This is particularly true in some of the rollings of flat stock where we pass from bar, to strip, sheet, or possibly a plate classification with only a very slight dimensional change. Each of these classifications of product have a distinct group of tolerances which are quite different from each other. The mill condition in most of these groups of sizes is approximately the same although tolerance is quite different.

To better explain this, chart number 10 shows a portion of a flat rolling schedule on our number two mill. The chart shows the size and classification, the standard tolerance, number of tons scheduled to roll, time required to complete the size including minor adjustments, and the number of billets required to fill the orders. Also shown in the column on the extreme right is the number of points which were put on the control chart when this schedule was rolled.

This I believe will give you some idea of our rates of rolling and the size of orders, and with our Quality Control station some 10 to 20 tons or more behind the mill, you can readily see that we are very often finished rolling the size before inspection has had a chance to check the section. Also that with our sample frequency set at 30 minute intervals, in many cases we do not get any measurements for the control chart. We must assume that control has been maintained for the entire period between samples.

Chart 11 shows a portion of the rolling of the schedule shown in chart 10. The chart shows the modified control limits jumping back and forth as the classification of material changes. On rollings of this type we take the narrowest tolerance for the group of sizes in the mill and impose corresponding limits for all sizes in that group. The objection raised to this type of a chart is, of course, that we are restricting the limits and that much more time than is necessary might be spent in maintaining the close limits. This, however is an educational problem and when proper discretion is used before holding for further inspection or slowing up production looking for trouble, the objection is easily overcome.

Imposing the tightest limits on this chart number 11 which with modified limits had only 2 points actually out of control would show 11 points out of control. Still the material represented by the original two points would be all that would require an investigation. This type of a chart has considerable in its favor but it must be understood by inspection and mill operators responsible for size control and production in order that production will not suffer unnecessarily.

Another study and application of product weight control although not having nearly the effect that size has is the control of length. Early studies made in December 1949 indicated approximately 100% out of control on a similar modified limit. Much work was done revising shrinkage and bogey factors. Also shear tables and stoppers were inspected and repaired to remove all possible play. After several months we had reduced this out of control material to 43.5%. This looked for a time as though it was about as far as we could go but after again studying our shrinkage and determining shearing temperatures a nomogram based on theoretical values was developed. This was studied and applied and we found another drop in out of control material. We now however ran into some short bars which could not be tolerated and in order to guard against this the control limits were raised so that the upper control limit and the tolerance coincided. This of course allows for some bars being in control but still slightly over specification. When this nomogram was put into use and inspectors supplied with a method for determining exact shearing temperatures the percentage of samples charted which were out of control droped to slightly less than 25%.

Shortly after our studies were made known and checks made elsewhere the carbon bar committee of the A.I.S.I. recommended an increases in tolerance. With the new increased tolerance we are ablt to keep this percentage of samples out of control at less than 10% in regular production.

In summarizing our results and with checks made in one of our customer plants where bar counts and inventories are accurately kept we are informed that we had reduced our amount of overshipment based on theoretical weight by approximately 1.00 % since the later part of 1948 and early 1949. Our figures for 1948 and the first few months of 1949 are limited but from those checks we have been able to make this figure does not seem hard to justify. It is calculated that the savings to this customer by our product weight control amounts to approximately $50,000 per year over our early 1949 level of operation.

Looking back over these changes in pass sizes and charts of past rollings our records indicate that in our number one mill we have reduced the size of our round and square products by .003 - .004 inches. In our number two mill an average reduction on these products of from .006 to .007 inches has been accomplished. Number four mill shows a reduction in variation from nominal size of approximately .003 inches.

With these figures in mind it can be noted that the weight per foot of the bars produced under Product Weight Control has been considerably reduced. The advantages gained by this closer control of product weight does not however have its greatest benefit at the steel plant because of our products being sold by weight. In the various customer plants however many benefits are available. Under a strict material control plan a more uniform number of pieces are available per ton of bars, which allows closer inventory control and also a closer to theoretical weight on machines built which contain large numbers of parts made from stock as received from the mill. On machined or forged parts less and a more uniform amount of stock must be removed by turning or flashing. This allows for extra tool life, more machine time, less freight on bars received, and also less expense due to scrap handling.

In order to show what this means in weight to our customers a comparison of charts 1 and chart 12 which represent an early and a recent performance on 1-3/16 inch round from number one mill can be made.

1949

Size
Average max +.0080" } +.0034"
Average min -.0012" }
Mean diameter of product = 1.1875 + .0034 = 1.1909"
Weight per foot = 3.7872 lbs
A twelve foot bar weighs = 45.4464 lbs
143 bars 12'-0" long weigh 6500 lbs

1952

Size
Average max +.0036" } -.0008"
Average min -.0052" }
Mean diameter of product = 1.1875 - .0008 = 1.1867"
Weight per foot = 3.7605 lbs
A twelve foot bar weighs = 45.126 lbs
144 bars 12'-0" long weigh 6500 lbs

Had we rolled our entire 1952 production of 710,000 net tons of bars into 1-3/16" rounds 12'-0" long we would have an equivalent of 221,850 additional 12'-0" bars. This in effect is a net gain of 5,000 net tons of steel which did not have to be, cut or flashed from bars, or did not find its way into excess weight of a finished product.

This provides a summation of the things we have done to date to control the weight of our product. We are continuing however with much of the work outlined and are expanding as fast as possible to cover all sections rolled. It is our intention in the near future to cover both those products sold on a dimensional tolerance and those special shapes and structural sections where the tolerance adhered to is on a weight per unit length basis.

CHART 1
3"
16
AVERAGE
RANGE
X.001
NO. 1 MILL
1949

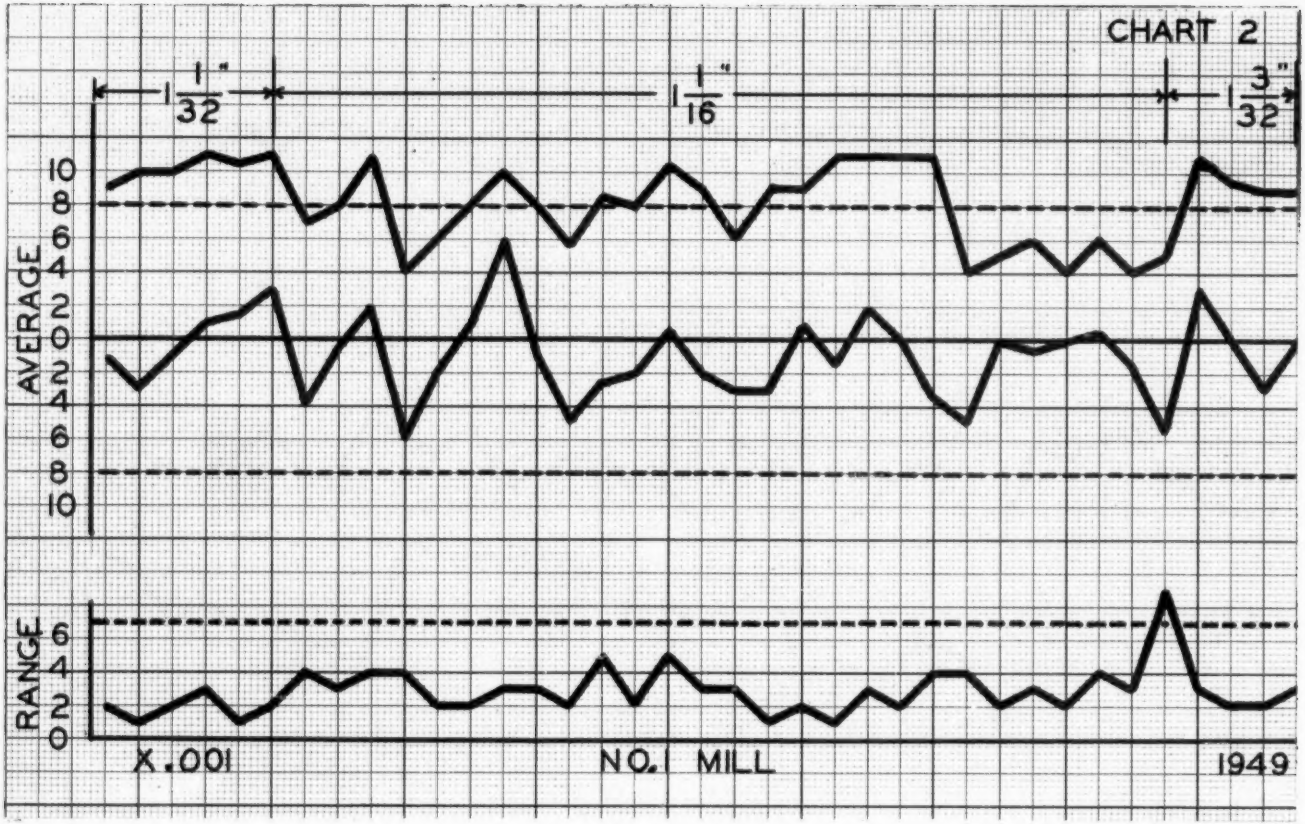


CHART 2
1 1/32"
1 1/16"
1 3/32"
AVERAGE
10
8
6
4
2
0
2
4
6
8
10
RANGE
6
4
2
0
X.001
NO. 1 MILL
1949

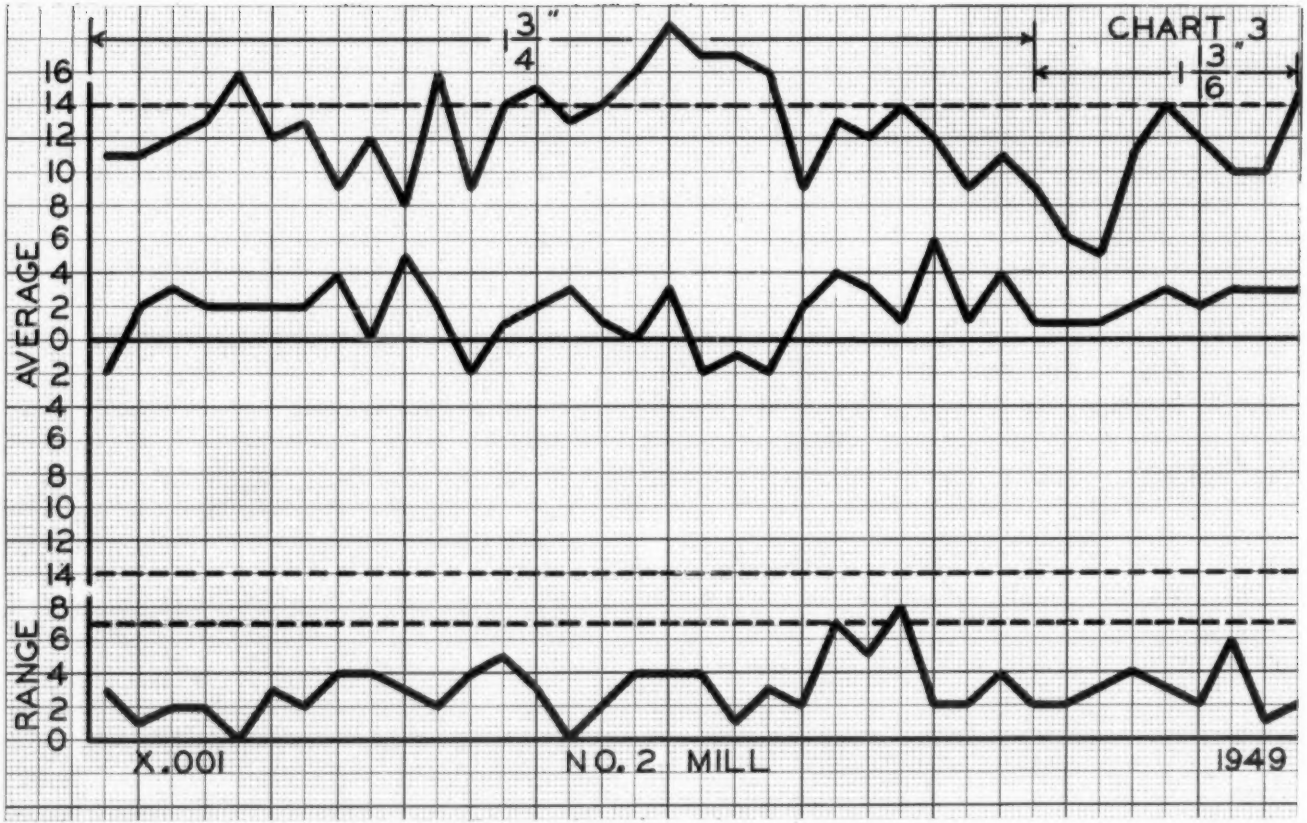


CHART 3
1 3/4"
1 3/6"
AVERAGE
16
14
12
10
8
6
4
2
0
2
4
6
8
10
12
14
RANGE
8
6
4
2
0
X.001
NO. 2 MILL
1949

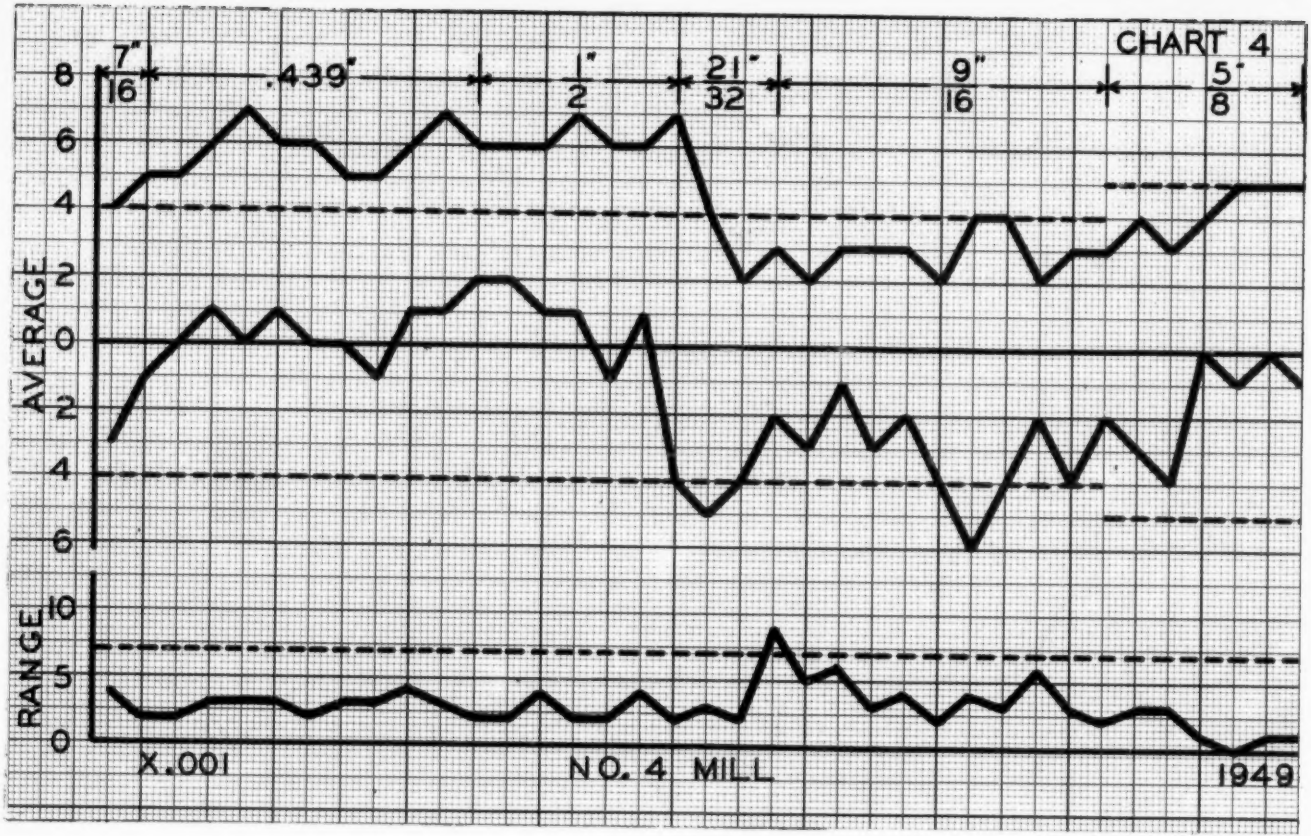


CHART 4
7" 16
.439"
1" 2
21" 32
9" 16
5" 8
8
6
4
2
0
2
4
6
AVERAGE
10
5
0
RANGE
X.001
NO. 4 MILL
1949

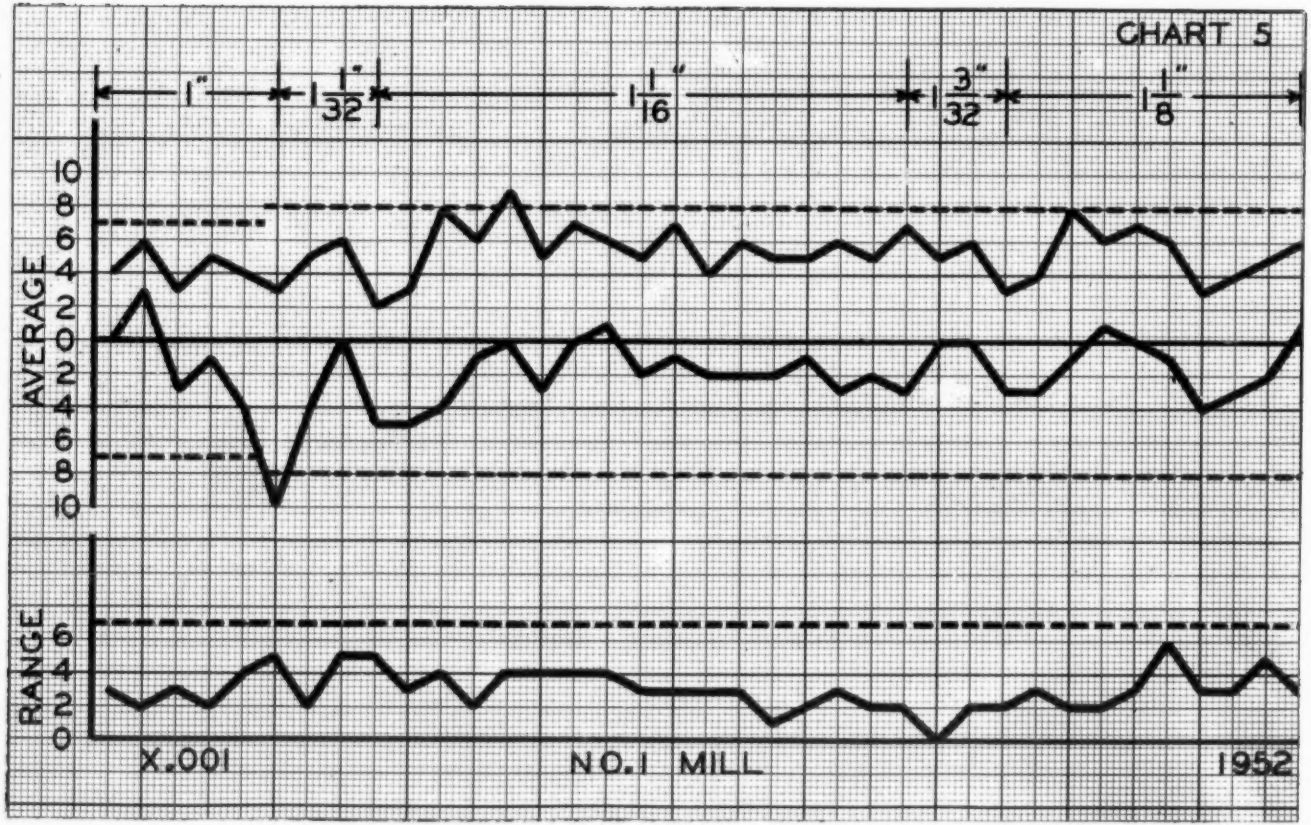


CHART 5
1"
1 1/32"
1 1/16"
1 3/32"
1 1/8"
AVERAGE
10
8
6
4
2
0
2
4
6
8
10
RANGE
0
2
4
6
X.001
NO.1 MILL
1952

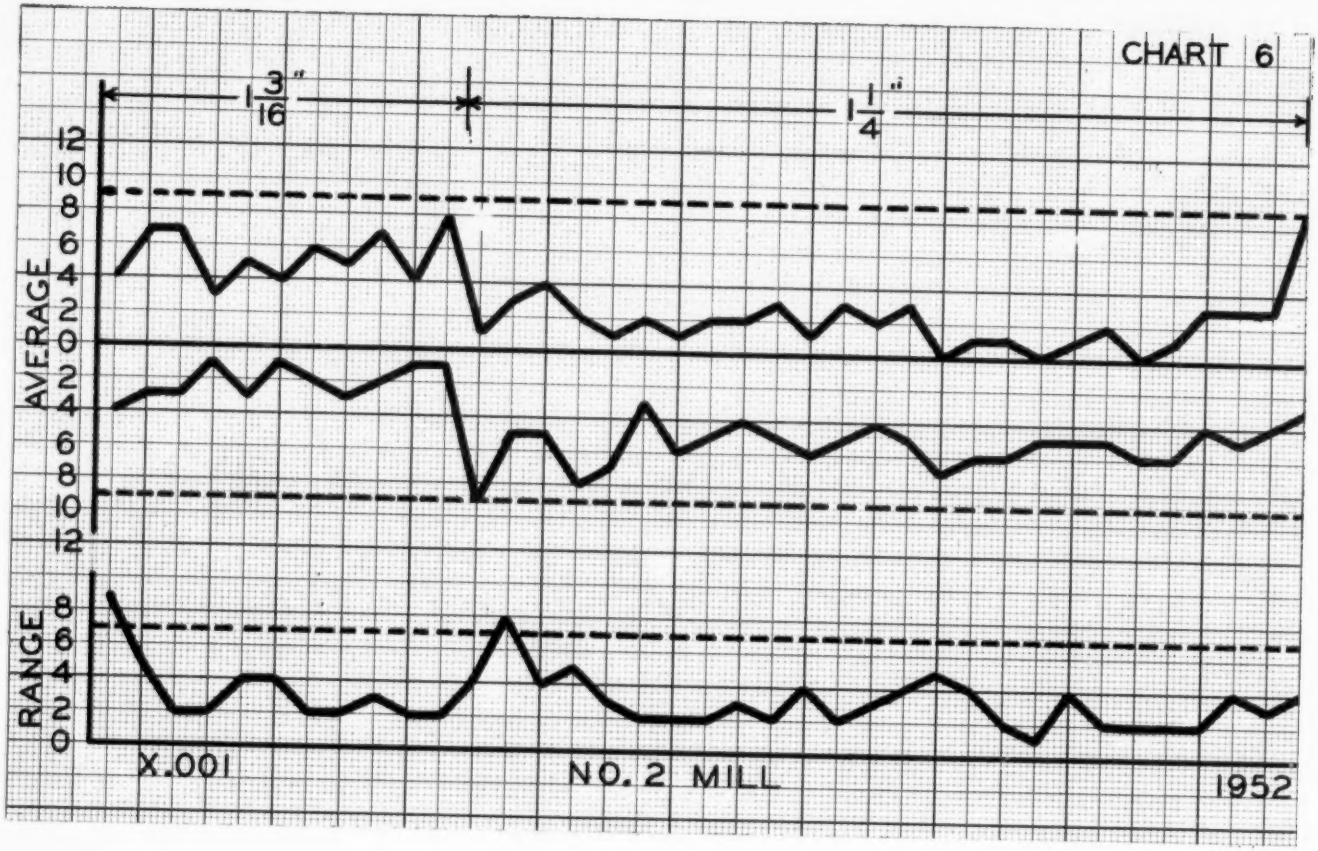


CHART 6
1 3/16"
1 1/4"
AVERAGE
12
10
8
6
4
2
0
2
4
6
8
10
12
RANGE
8
6
4
2
0
X.001
NO. 2 MILL
1952

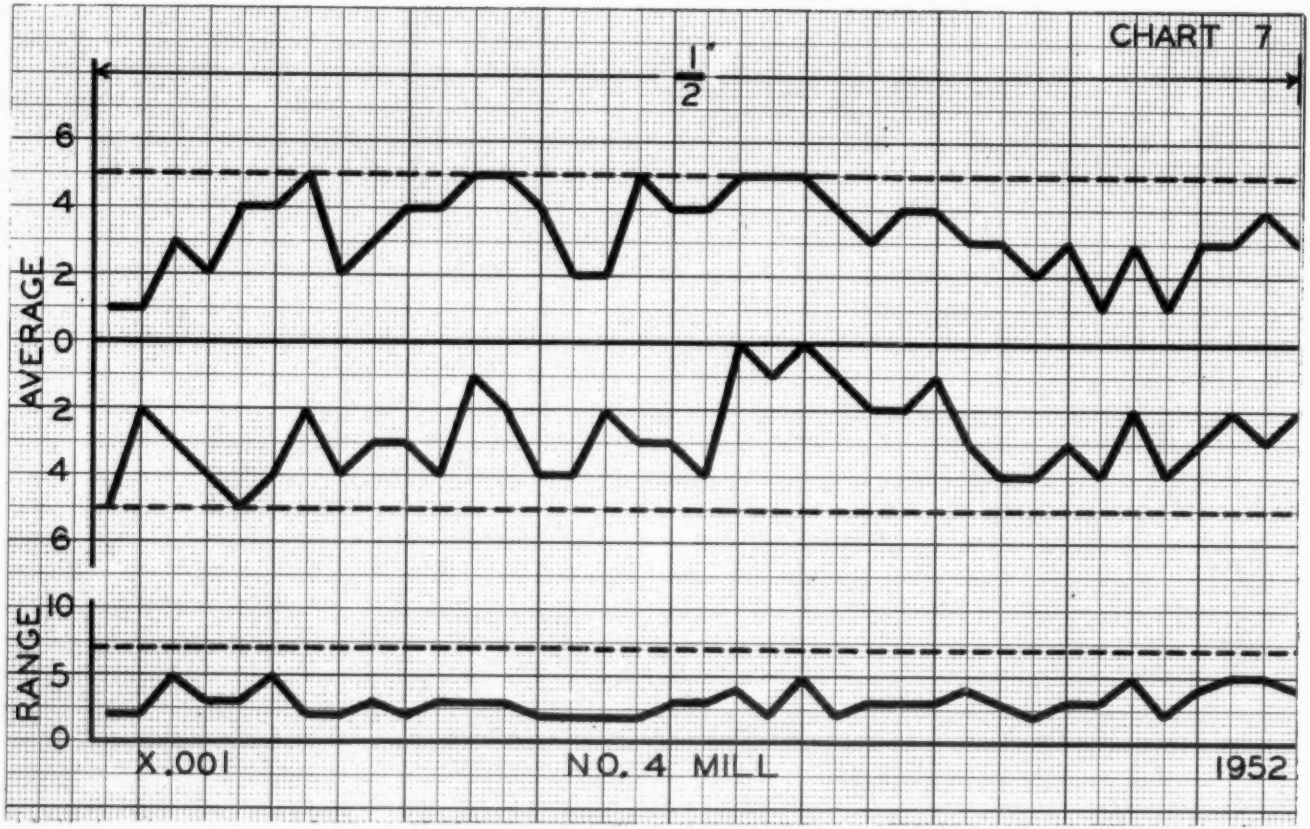


CHART 7
1/2"
6
4
2
0
2
4
6
AVERAGE
10
5
0
RANGE
X.001
NO. 4 MILL
1952

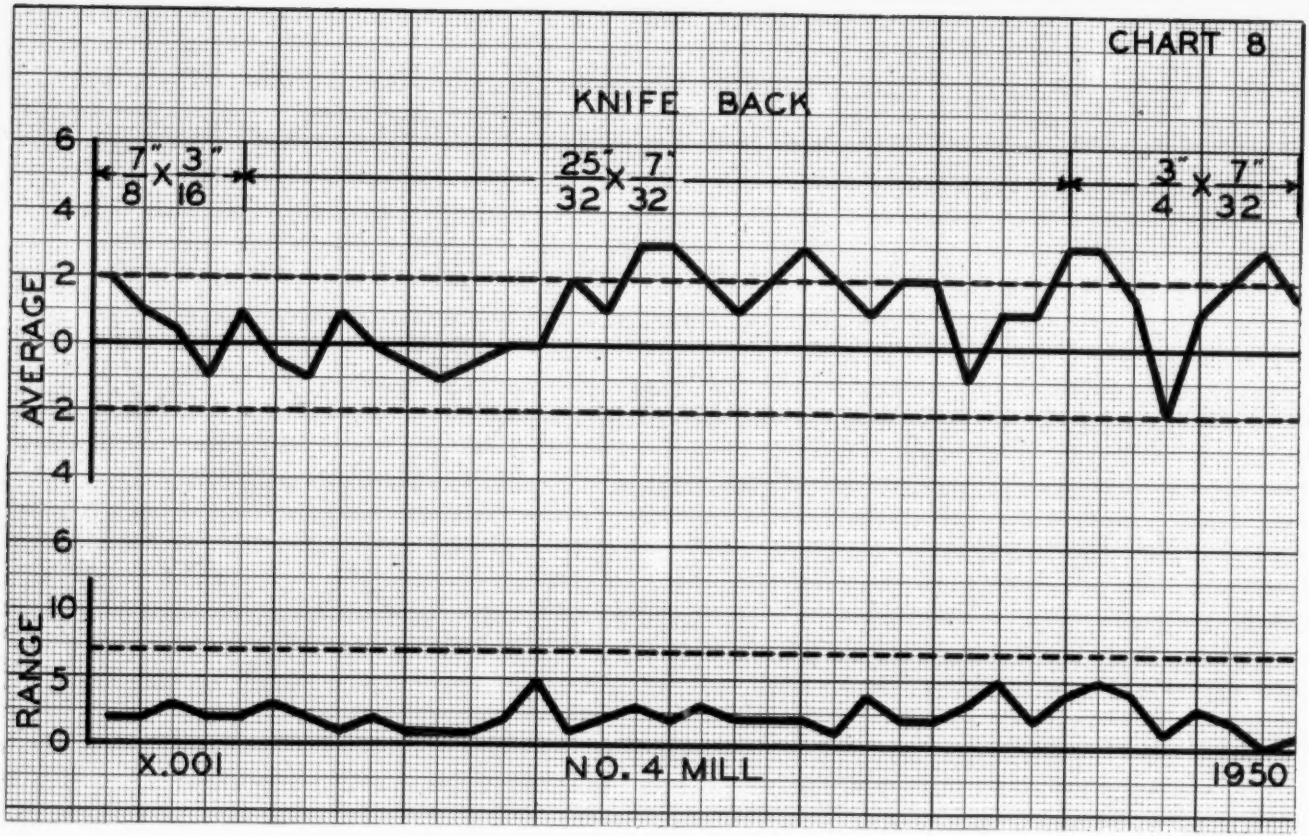


CHART 8
KNIFE BACK
7" X 3"
8 16
25" X 7"
32 32
3" X 7"
4 32
AVERAGE
6
4
2
0
2
4
RANGE
6
10
5
0
X.001
NO. 4 MILL
1950

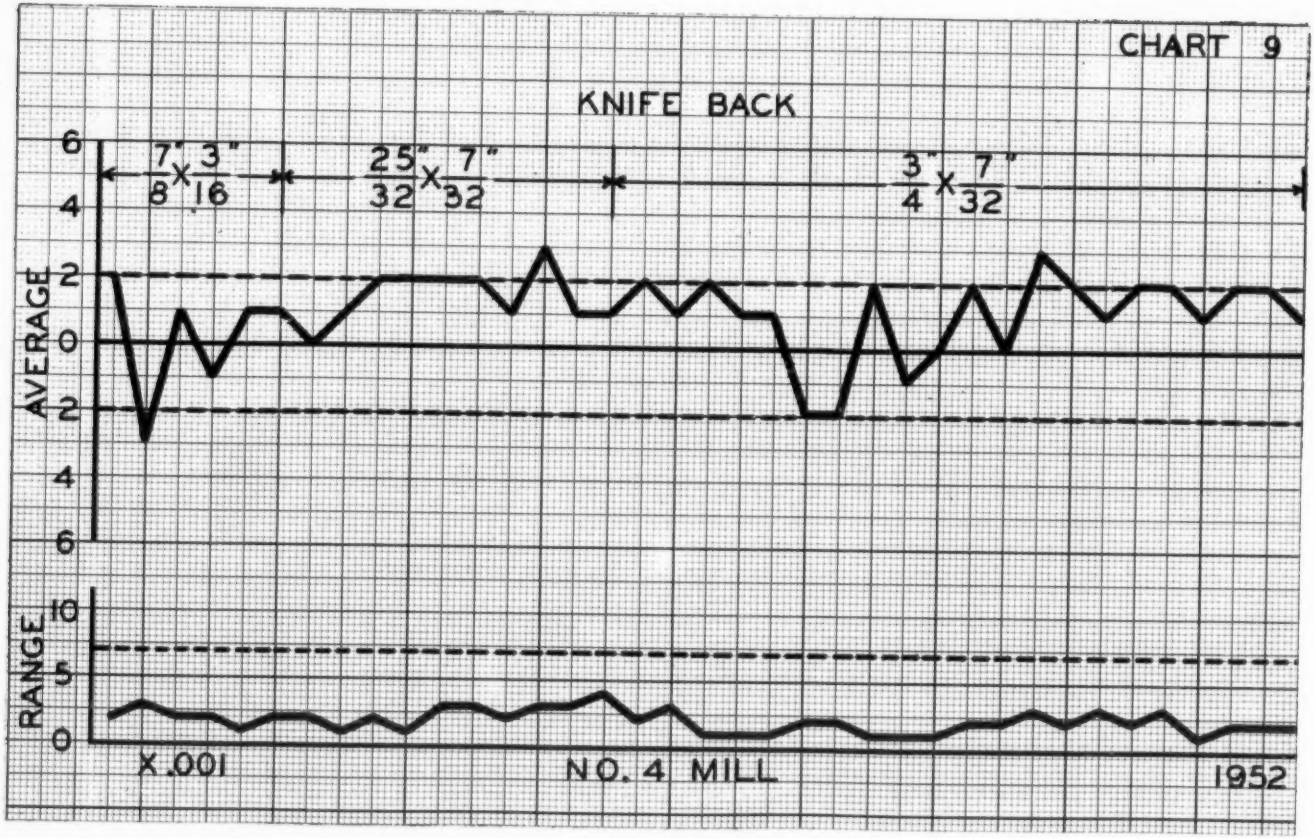


CHART 9
KNIFE BACK
7/8" X 3/16"
25/32" X 7/32"
3/4" X 7/32"
AVERAGE
6
4
2
0
2
4
6
RANGE
10
5
0
X.001
NO. 4 MILL
1952

CHART 10

SIZE & CLASS	STD. TOL.	TONS SCHD.	ROLL TIME	NO. PCS.	PTS. CHT.
6 x 1/8 S	± .005	30	1:15	298	3
5-7/8 x 1/4 F	± .015	3	:10	26	1
5-13/16 x 13/64 F	± .009	45	1:50	320	4
5-3/4 x 3/16 S	± .006	12	:26	94	1
5-3/4 x 1/4 F	± .015	30	:24	96	1
5-1/2 x 1/4 F	± .015	5	:12	40	1
5-3/8 x 1/4 F	± .015	6	:18	48	1
5-1/2 x 3/16 S	± .006	6	:13	46	-
5-1/2 x .187 S	± .005	7	:15	64	-
5-1/2 x 1/8 S	± .005	6	:15	66	1
5-1/4 x 1/8 S	± .005	3	:12	40	-
5-1/4 x 3/16 S	± .006	43	1:32	400	3
5-1/8 x 5/32 S	± .005	3	:15	28	1
5 x 1/4 F	± .015	223	5:28	1668	10
5 x 3/16 S	± .006	12	:20	94	1
5 x .156 S	± .005	9	:25	84	1
4-3/4 x 5/32 S	± .005	3	:10	36	-
4-3/4 x 1/4 F	± .015	3	:08	22	2
4-1/2 x 9/32 F	± .015	10	:32	102	1
4-1/2 x 1/4 F.	± .015	9	:08	78	1
4-1/2 x 3/16 S	± .006	21	1:27	196	-
4-1/2 x 1/8 S	± .005	3	:10	45	1
4-3/8 x .1793 S	± .005	3	:07	32	1
4-1/4 x 1/4 F	± .015	3	:13	32	1

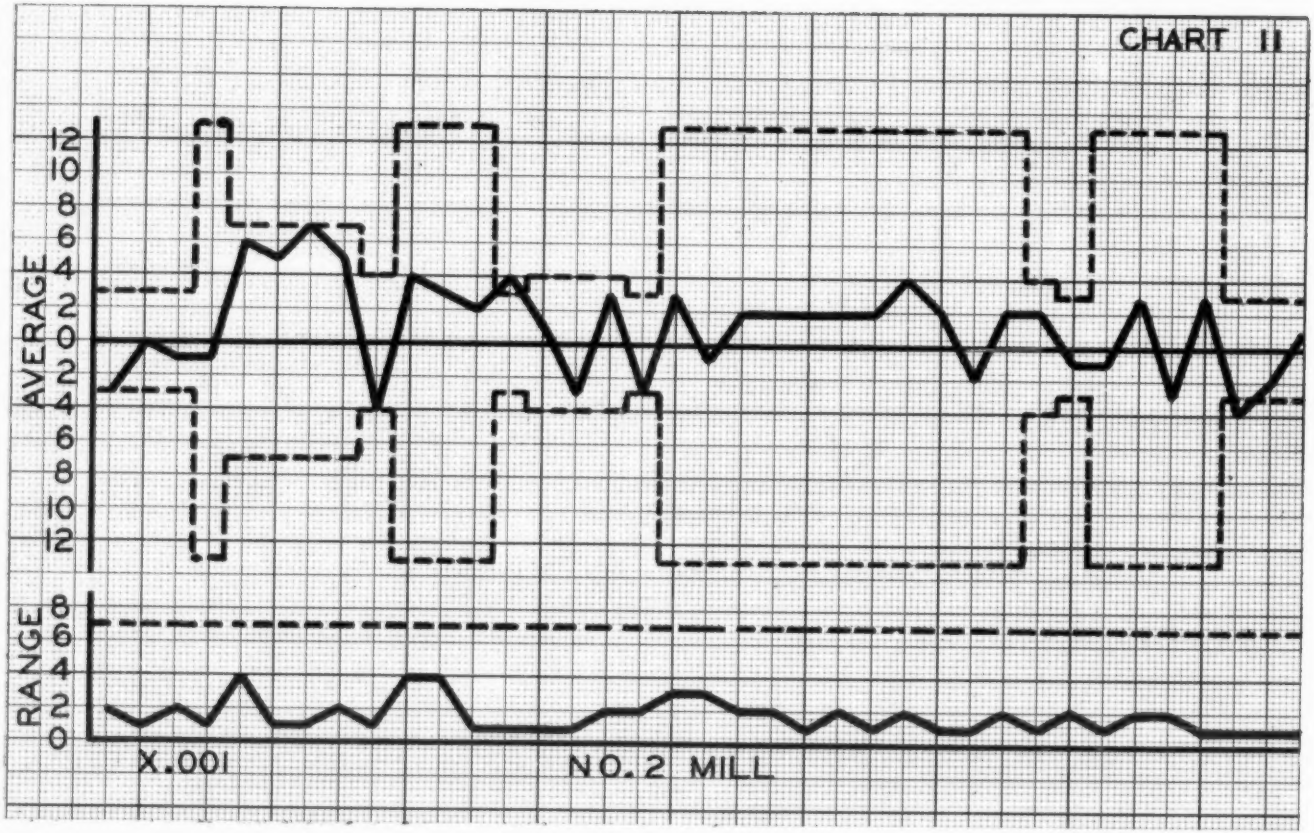


CHART 11
AVERAGE
12 10 8 6 4 2 0 2 4 6 8 10 12
RANGE
8 6 4 2 0
X.001
NO. 2 MILL

CHART 12
3 1/16
AVERAGE
RANGE
X.001
NO. 1 MILL
1952

QUALITY CONTROL OF VISUAL CHARACTERISTICS IN CARPET MANUFACTURE

Albert G. Klock
Bigelow-Sanford Carpet Company, Inc.

One usually thinks of Statistical Quality Control as applied to dimensional characteristics. There have been few statistical applications of those characteristics which cannot be measured. Most products are made to dimensional specifications and yet have appearance requirements - such is the product of the carpet industry. Our claims showed a real need for some means of measuring visual characteristics, if we were ever going to know where to peg our Quality level.

The method to be described here is not an accurate or fool-proof means, but rather represents a workable approach. The carpet industry has two classifications for its product - perfect and mill second. Mill second carries a discount. The production cost is often more for mill seconds than for perfects due to the additional work required in attempting repairs. The discounts and additional production costs are often enough to mean the difference between profit and loss.

Our problem of classification of carpet is to use it consistently as the average customer. The secret here is to conduct the inspection in the same mental sequence as the customer. This principle consists of grouping visual defects into basic classes of visual characteristics, which we have developed to fit the carpet needs, namely:

a. Color - uniformity or levelness
b. Yarn - physical differences which are visible
c. Weave - machine difficulties, carelessness in assembling
d. Finish- feel, handle, backsizing, off-square, dressing up of a poor appearing fabric.

Our next requirement is the attachment of a numerical value to each of the basic classes. Our objective is to keep the readily apparent visual defects from getting to the customer. On that basis, we assign the values to the most outstanding defect in the basic classification. We have found that the normal range of opinions is in five steps. By that we mean if an individual were to try to describe the range of opinions of Quality in a product, he would most likely say that the product is:

Best - perfect - completely absent of flaws, exceptionally desirable

Good - contains a few minor flaws which are not readily detectable and will not effect or harm the usefulness of the carpet - still a very desirable product

Satisfactory - readily detectable flaws of a minor nature, no damage to the end use - satisfactory service or use could be expected

Chance - readily detectable flaws of a minor nature and even of a major defect that the average customer might not accept

Inferior - contains major defects which cannot be repaired and must be rejected - no customer will accept.

We merely assign numbers to the degrees of opinions as follows:

0 - Best
1 - Good
2 - Satisfactory
3 - Chance
4 - Inferior

Once this transition takes place, the whole idea of opinion rating of Quality for visual defects takes on the numerical aspect. To assist the inspector, we have developed guides for judging the basic classifications by listing all of the physical defects possible and each of the basic classifications. This guide is in a chart form, with a possible degree listed for each defect. In accordance with the Inspector's Guide, the inspector notes his ratings on a daily report form which gives identity to the product inspected. A summary of this daily report gives the average rating values of the major characteristics and also a daily Quality level. This daily report is combined to weekly, monthly, quarterly and yearly summaries and is a valuable tool to detect trends and to point out trouble areas.

To assure the correct sampling, sets of random sampling plans have been devised which the inspector follows in the selection of his samples. This predetermined sampling plan gives us a good cross section of our product.

After we have had this plan in operation for some time, we desire to find out how well our inspectors would repeat themselves, and thereby determine the validity of this inspection. The question of repeatability concerns itself with those factors which contributes most heavily in the variation in measurements, namely:

1. The variation within inspectors
2. The variation between inspectors
3. The variation between shifts - day versus night

Our method here was to select ten rolls of assorted grades in various Quality levels and have the inspector evaluate them at four different times, so spaced that the ratings would not be remembered. They were reviewed twice under daylight conditions and twice under night or fluorescent lighting conditions.

Using the Student "t" for pair groupings, we found at the 5% significant level the following:

1. The judgment of each inspector on repeated ratings is consistent.

2. As to variation between inspectors, we found the significant difference in the weave category. This was corrected by better definitions of the visual defects in this classification.

3. There is a significant difference in the results between daylight and nightlight inspection in all categories except yarn. Yarn is the most physically measurable of the four classifications and, therefore, the most consistent. This pointed out the need for standardization of lighting conditions under which the inspection of carpet takes place. Much has been done in this area.

All surveys and checks on this system so far have given us assurance that we are doing something constructive. We are gathering and dispensing information which allows those truly responsible for Quality and production people to concentrate their efforts in those areas where weaknesses become quickly evident under this plan.

The following points are the values of such a visual inspection plan:

1. The inspection serves as an audit on production employees and other inspectors when this plan is used as a check inspection on production inspectors.

2. A single index of Quality referred to as a Quality level is obtained for each division or basic operation.

3. Comparable indices are developed for each individual grade and thus Quality levels and characteristics or standards can be developed for grades.

4. The advantage of tieing in the functional Manufacturing Departments quality-wise with the end production.

5. Assurance to top management who is vitally interested in the Quality of outgoing products.

6. The assistance of efficient handling of claims to the records maintained by this system.

7. The improved consistency of inspection results which can be obtained through a systematic approach through the use of opinions.

8. A reduced time required to train the inspector through complete definitions of defects and a formalized plan for training.

9. A yardstick by which to measure competitors goods.

10. The advantage of theoretical point of view where we have changed our inspection from attributes to variables plan.

In summary, here is a refined technique for handling Quality problems relating to products which are manufactured to physical specifications and sold on the basis of their visual characteristics. The working details have been described, and the more obvious advantages emmerated. They are not entiraly perfected. It is of great values in aiding a control of the appearance of product by means of opinions.

INDUSTRIAL APPLICATIONS OF OPERATIONS RESEARCH

John B. Lathrop
Arthur D. Little, Inc.

Most of you are here because of your interest in the development of a scientific basis for controlling the quality of industrial production. You may also be interested, then, in operations research, which applies the methods of scientific research to control and decision problems in many areas of business and industrial management.

This discussion will outline briefly how operations research began, what it is and how it works, with some examples of its use. We will also touch upon what it adds to present management techniques, and some recent developments in the field.

How Operations Research Began. Organized operations research was a war baby, and commenced with the formation of groups of scientists to analyze combat data in wartime, in order to determine the effectiveness of selected military operations, and to devise means for improving the use of weapons and weapons systems. The techniques of scientific research were applied to many operating problems never before approached in this manner, many of which were widely held to be unsusceptible to scientific or quantitative methods of attack.

In anti-submarine warfare, in connection with the terrific problem of finding enemy submarines at sea, OR isolated and measured from combat records the effects of searcher speed, searcher altitude, and meteorological conditions on the probability that a submarine would be sighted. Physiologists contributed factors for the behavior of the eye, electronics engineers factors for the behavior of radar--to be worked into an overall theory of the search operation. The theory, in turn, provided the basis for best decisions as to allocation of searching effort, and as to which equipments were best.

Similar theories, or working models of the operations, provided tools for the best decisions regarding means for anti-submarine attack, for arranging convoys and their escorts, for bombing factories, for gunfire support of amphibious operations.

Operations research achieved stature in the Armed Forces when almost anything new was tried at least once in the desperate effort to defeat our enemies. Many new methods were called, but few were chosen. Operations research stayed, and has since become an important and accepted feature of our military organization, a tribute to its value and usefulness.

OR emerged from the war to find peacetime battles of production, sales, and distribution looking like familiar ground. Here, too, problems outside the conventional fields of science and engineering have yielded to the methods of scientific research. It has been found that many of the principles developed from military operations research could be transferred directly to the problems of business and industry.

What It Is. Many management decisions in industry require the selection of factors relevant to the outcome of an operation, determining values for each of these factors, estimating how each affects the outcome of the operation concerned, and, finally, selecting a course of action

which will yield the most favorable outcome. Too often such decisions are based upon an incomplete understanding of the basic relationships among important factors, particularly when such relationships may be obscured by masses of unwieldy data. The result may be judgments based in too large a part on opinion, so that conflict and disagreement result.

Operations research tries to provide a more complete and integrated quantitative basis for such decisions, by defining explicitly the factors which should be measured; finding ways to isolate them from among complex data and to determine their values; and establishing explicit quantitative relationships among them which can be used to predict the effects of alternative decisions.

An operations research attack on a problem is basically that of the scientific method, an unbiased systematic approach. Many of the techniques used are not new, only their wider applications. Scientists do not know more about war than military men, or more about business than businessmen, but they do bring to their problems a new attitude, as well as new techniques, and formulate problems in a new language.

How It Works. The operations research approach may commence with the selection of measures of effectiveness in terms of which the success or failure of an operation is to be judged. These may be quite different from those previously employed. As an example, during the war the Navy provided crews to man anti-aircraft machine-gun defenses for merchant vessels. It was noted that the proportion of attacking enemy aircraft shot down by such batteries was very small compared with AA experience of naval warships. It was suggested, therefore, that the naval gun crews could find better employment elsewhere and should be withdrawn from merchant vessel assignment. Closer study indicated, however, that in the case of merchant vessels defended by machine-gun batteries, the ships hit per attacking aircraft were much lower than for undefended vessels. The correct measure of effectiveness to employ in this decision was ships hit per attacking aircraft, rather than aircraft shot down per attack.

A recent operations research study concerned with evaluating the effectiveness of a chain of retail outlets for a large manufacturing company found that different officers of the company were, perhaps unconsciously, applying widely differing measures of effectiveness in reaching conclusions as to the success of the retail chain. (In this, and following examples, we are forced to paraphrase or abbreviate statement of the problems, in order to protect the confidences of the companies involved.) The president felt that the chain should be operated in such a manner as to maximize profits to the company; the treasurer wanted to see them adjusted so as to maximize return on invested capital; the sales manager was interested in gross margin, or sales less sales expense. It was discovered that these varying measures of effectiveness were mutually incompatible, and that the store chain could not be operated to maximize them all at once. It was necessary to develop a new and different measure which represented in proper proportions the interests of the various officers of the company, and to which they could all agree.

The next step in the operations research approach may be the development of a working model of the operation studied. This model usually takes the form of a set of mathematical expressions, since words are

unwieldy, and because mathematical equations provide means for manipulation, extrapolation, and maximizing. The model provides a theory of the operation, which can be used to predict quantitatively the effects of alternative courses of action. It is not an abstraction, and every part of it should have a physical and practical meaning.

In order to illustrate what we mean by a working model, let us look at a simple theoretical example of the development of a model of an operation. In a certain production machine is a bearing secured by two bolts, which are subject to breakage. If as many as one of these bolts is intact, the machine can be operated. If both break, however, the machine is down for twenty hours for repair and adjustment. An inspection of the machine while it is still in running order, to ascertain the condition of these bolts, and to replace one if broken, would require the machine to be down for two hours. The decision required is whether or not to institute a program of regular inspections of these bolts and, if so, the interval at which such inspection should be carried out.

The measure of effectiveness selected is the expected fraction of the time that the machine is out of service, which we wish to minimize. An examination of operating data showed that the average lifetime of a bolt is 1,000 operating hours, provided that both are intact and sharing the load. After one bolt is broken, the average lifetime of the remaining bolt was found to be 200 hours. Furthermore, an examination of the distribution of bolt lifetimes around the average showed that breaks appeared to occur at random, so that distribution of bolt lifetimes could be taken to be exponential.

From this information and probability theory, it is simple to produce a model of the operation in the form of probabilities that none, one, or two of the bolts would be broken after a machine had been operated for any given number of hours after an inspection. These probabilities would look about as in Figure 1.

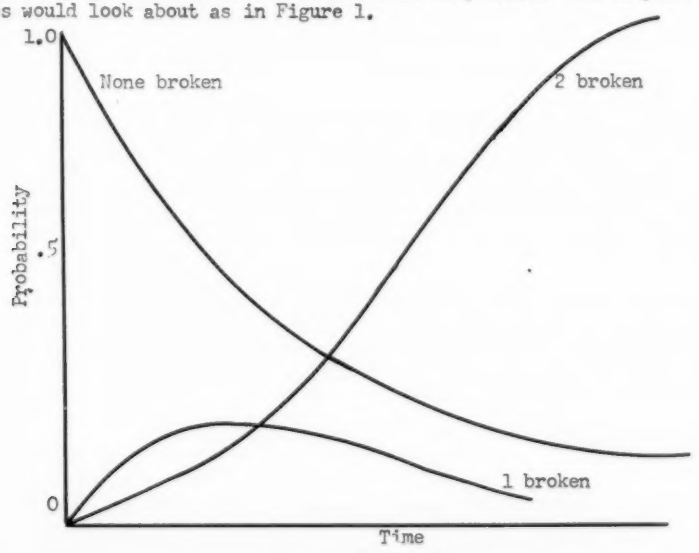


Figure 1

An extension of this model to include the times required for inspection or repair provides a relationship between the fraction of the time that a machine should be out of service and any given inspection interval, which takes the form of Figure 2.

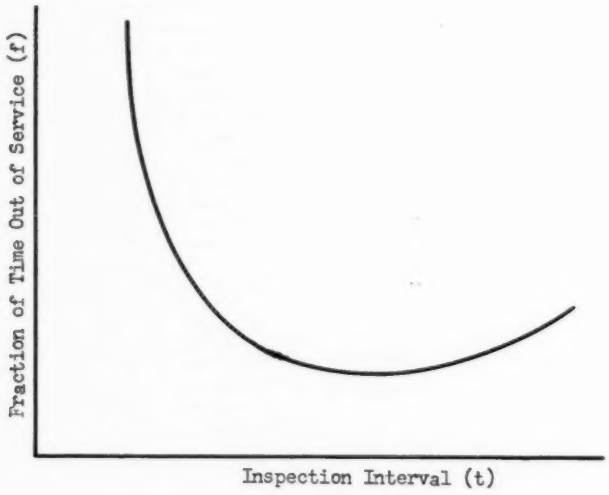


Figure 2

For those interested, the formula for this "decision function" is:

$$f = \frac{20 - 30e^{-t/500} + 12e^{-t/200}}{720 - \frac{2590}{3}e^{-t/500} + \frac{436}{3}e^{-t/200}}$$

An expression like this, rather than something constructed from wood and steel, is the "working model" as seen by operations research. In this simple illustration the techniques needed to develop it were relatively simple. In practice, techniques beyond those normally used in engineering--such as communications theory, game theory, linear programming, or variational calculus--as well as probability and statistical theory, are often drawn upon.

Another step in the operations research approach is typically to check the working model against operating or experimental data reflecting varying conditions, in order to determine whether the model can be counted upon to reproduce actual results when such factors as time, place, or type of product are varied; adjusting the model if need be.

As a final step, the adjusted model is used to provide a useable basis for decision by solving it for the optimum levels at which the operation should be carried out, predicting the effects of alternatives, and devising simple rules or control charts so that operating personnel can apply the rules for best operation.

Example - Evaluation of A New Development. In connection with the development of a new transportation control equipment, a company required a quantitative measure of the degree to which it would improve the operation. The technical characteristics, costs, and reliabilities of the proposed equipment and those which it might replace were pretty well known. There was no good way, however, to place it in an operational rather than a technical environment, and there to compare it quantitatively with competing equipments.

A principal contribution of OR to this problem was to develop a new measure to describe an elusive characteristic of railroad operations--the delay to trains caused by one train waiting for another at crossings or sidings--akin to a safety factor. Neither the developing company nor the operations research team were prepared to tell railroads what this time delay factor should be and practice varied widely. But, they could provide a measure for it, and show how it affected the control operation.

So a measure called a "passing parameter" was developed. This was merely:

$$P_o = \frac{T_1 - T_2}{t}$$

where T_1 and T_2 were the times at which trains were scheduled to arrive at a possible point of conflict--such as a siding on a single track road--and t was the time it would take the faster or heavier train to decelerate and stop.

A working model for a typical actual stretch of road was also developed, in terms of the speeds, accelerations and densities of trains thereon during a typical operating day, and the number of meeting points involved. The measure of effectiveness chosen to compare alternative control systems was total train delay per day due to control. From these measures a quantitative picture of the relative operational suitability of alternative control systems could be presented in terms of a factor which remained under the control of the operating personnel.

Example - Production and Inventory Scheduling. A problem frequently encountered in one form or another is how to schedule production and inventories to meet seasonal or severely fluctuating sales, at minimum overall cost. In a recent case, existing policy was to schedule production relatively uniformly throughout the year to meet the sales forecast with a minimum of overtime and alteration of labor force, letting inventories take up the slack.

Although this suited the production manager, the treasurer was unhappy and thought this practice involved unnecessary storage costs and tied up funds in too large inventories. The sales manager was satisfied as long as inventories were always large enough so he could meet demand without running into back orders and their attendant loss of good will, loss of customers, and artificial stockpiling. Scheduling had been difficult because of the large number of products and production lines, and several factors were involved which weren't measured and could only be argued about.

The working model of the operation used the total of costs associated with the schedule as a measure of effectiveness to be minimized, and summed separately the costs associated with production, inventorying, and

back-orders--each group representing the interests of one of the interested executives. Severely summarized, it looked like this:

$$\text{Cost} = \sum C_P(P_m) + \sum C_I(I_m) + \sum C_{BO}(I_m)$$

where:

$C_P(P_m)$ = Cost of producing (in any month) an amount of production P_m

$C_I(I_m)$ = Cost of carrying in inventory an amount of inventory I_m

$C_{BO}(I_m)$ = Cost of back-orders associated with an amount of inventory I_m

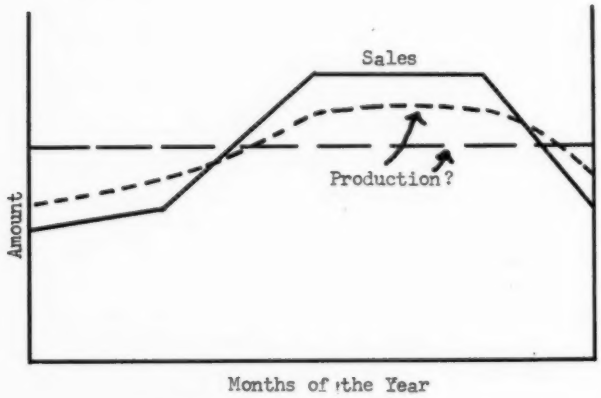


Figure 3

Once the model was set up, the best schedule can be solved for by the technique of linear programming, an arithmetic for solving such an expression, subject to restrictions imposed by capacities to produce or store, and company policies such as those relating to permissible overtime.

Important features of the operations research problem, however, were the limiting of costs appearing in the working model to those related to the production schedule--when the goods were produced--and assigning values to some of them which weren't measured before. For instance, the cost of raw materials had to be borne sometime during the year, so was essentially unaffected by the production schedule. On the other hand, costs of overtime production had to include costs of setting up extra shifts and any increased costs due to less efficient labor and supervision. Back-order costs, except for those due to extra handling and bookkeeping, had not been measured before. Although it was difficult to cost loss of customers and good will, solving the model against previous production schedules revealed the values unconsciously assigned to back-orders in practice, bringing them out into the open for examination. Graphical methods were developed to show the effect of the choice of back-order costs on minimum inventory levels and carrying costs.

The sum of overtime and inventory costs for the optimum schedule turned out to be 44 per cent of those for the schedule previously planned. A quick and simple method for extending optimum scheduling to all production divisions of the company was devised, as were methods for altering schedules with changes in costs. Guesswork in scheduling was eliminated and argument reduced to discussions concerning the values of component factors, with knowledge as to how variation in any of them should affect overall costs.

Example - Planning Machine Runs. A manufacturer of small components on a job basis was faced with the problem of estimating the number of pieces to produce on any one order, since the fraction of pieces discarded because of scrap or spoilage varied widely from run to run of the machines. Customers normally would accept up to a 10% overrun, but if larger numbers of acceptable pieces were made, the excess could only be scrapped. If, on the other hand, spoilage resulted in the number of acceptable pieces being less than ordered, an expensive new machine set-up and rerun was required.

The operations research team, after examination of production data, found that the distribution of finished output as a fraction of started quantity could be expressed by the equation

$$\phi(x) = k\, x^{a-1}(1-x)^{b-1}$$

where:

x = Fraction of started quantity which was acceptable.
$\phi(x)$ = Relative frequency of runs producing an acceptable fraction x.
k, a, b = Constants, to be solved for.

The analysts found that this expression fitted data for all products, with the constants k, a, and b varying somewhat among items, and, when charted, appeared as in Figure 4.

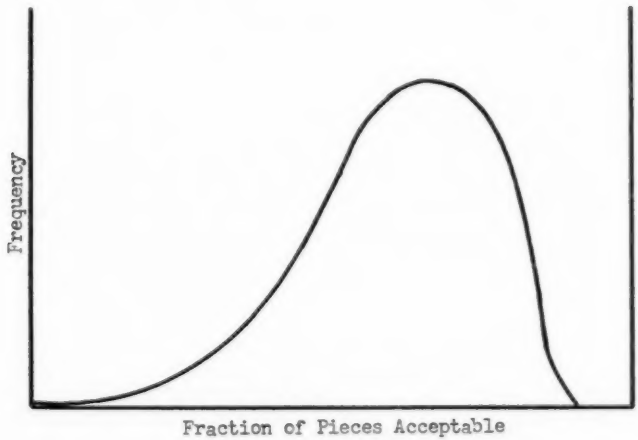


Figure 4

Since gross profit was the measure which the production manager wished to maximize, this was selected as the measure of effectiveness, and the following expression derived:

$$P = P' - s\int_{0}^{A/S} \phi(x)dx - cS\int_{1.1\,A/S}^{1} (x-1.1A/S)\phi(x)dx$$

where:

P = Gross profit
P' = Gross profit if just the right number of units is made
A = Number of finished pieces desired
S = Number of pieces started
s = Cost of an additional machine set-up
c = Cost per unit produced.

The second term of this formula represents the average loss due to additional set-up, and the third term the average loss due to overrun.

Maximizing this expression for profit yielded a curve such as that of Figure 5, from which to choose the optimum number of pieces for an initial run:

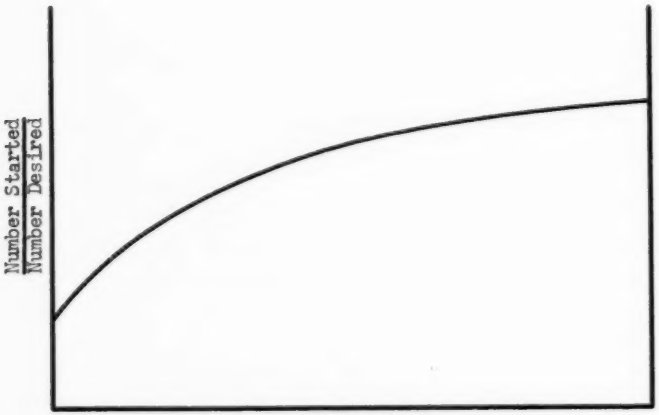


Figure 5

This curve shows, of course, that the larger the machine set-up cost the more leeway for waste should be provided in the number of units started. The proper number of units started could be readily computed to provide the optimum balance among additional set-up cost, surplus scrap cost, and order size.

Example - Determination of Production Cost Factors. Estimation of unit product costs is of prime importance to industrial managers as a basis for controlling the efficiency of production operations, as part of the information needed for setting pricing and sales policies, and as a guide to direction of engineering effort. The two most common sources of unit costs appear to be time studies and engineering or technological requirements or limitations.

There are many cases when time standards are neither specified by the technology nor readily determinable by time-study methods. The operating times of complex semi-automatic multi-purpose equipment with fairly long production cycles may depend both on the efficiency of the operator and, in a complex way, on a number of characteristics of the input materials. The length of the individual production cycle, the wide variation in operator tasks depending on the particular job assigned, and the number of possible uses of the equipment, may make usual time-study or job-analysis methods unworkably cumbersome. In such cases, operations research analysis of routine statistical data from shop operations in the light of a realistic mathematical model reflecting the important machine and product characteristics may be used to derive useable time standards.

An example arose recently in a study of some operations in an electrical wire plant. The operation concerned was that of "horizontal taping", in which a partly completed cable is covered with a helical layer of cloth or metal tape. This is done by pulling the cable through the center of a rotating disk, the "taping head", which carries a roll of tape. The tape is unwound, and through a set of rollers presented to a moving cable at an angle such that in one rotation of the taping head one turn of the helix is placed upon the cable with a prescribed (small) overlap.

There are thus two inputs: First, the cable, which has two important characteristics--the diameter, and the state of cure of the rubber insulation; second, the tape, which also has two characteristics--width, and kind of tape. Only a single product was produced from a given set of input materials, so that there was no dependence of production rate on the product characteristics.

In setting up the model for this machine, some assumption had to be made concerning the nature of the property that limits the speed of taping. It was assumed that taping speed is limited by the tensile strength of the tape, which must be unwound from its roll at a high rate. A model was established from the following geometrical characteristics:

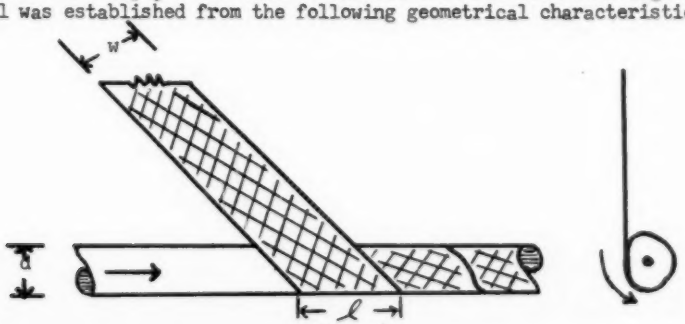


Figure 6

On the assumptions that the tension in the tape is proportional to its speed, that the strength of the tape is proportional to its width, and neglecting overlap, the limiting machine speed is:

$$K \frac{w}{\sqrt{\pi^2 d^2 + \ell^2}} = \frac{\ell \pi d}{\pi^2 d^2 + \ell^2}$$

and the time required to produce a length L of product is

$$T = t_o + K \frac{L}{\ell} \left(\frac{\pi d}{\ell} + \frac{\ell}{\pi d} \right)$$

When attempts were made to fit this equation to available data (time and payroll information for several hundred jobs of varying length, cable diameter and lay) values of the constants for the three machine types involved were inconsistent, and, in some cases, physically meaningless. However, by recognizing correlation among variables and bias caused by the prevalence of short jobs, modifying least-squares procedure by trying geometric means and combinations of variables, satisfactory constants were found. The model then explained 94% of the original variance with physically meaningful constants, and revealed the fundamental differences among machines and products.

It was possible to set up a job "quality control chart", based on estimates of the standard deviations of the values of t_o and K. There is considerable uncertainty in the time required for identical jobs, but with the aid of this chart, the foreman is able to decide if the time required is unreasonable or not.

Example - Evaluation of Advertising As a final example, a company sells a large number of products, all of which are normally promoted by direct mail and magazine advertising. Their advertising executive has a difficult problem in deciding how much of his department's available effort should be budgeted to each product.

In setting up such budgets he must take into account such factors as the total market potential of each product, differences between products, appeal to the consumers, and the rate at which sales will drop off if he stops advertising. He, of course, considers a great many other factors such as differences between media, and the extent to which competition is advertising its products.

In order to help him, an operations research group set about to see how much of the variations in the sales history of products could be explained by a model which takes into account the advertising effort devoted to each and three of the principal relevant factors: sales potential, a measure of the market response to advertising, and the rate at which sales would die off in absence of advertising. Other factors were assumed to average out over all products and were neglected in initial analysis.

The sales potential represented the amount of the product which the market could absorb with an infinite amount of advertising, the saturation level. The market response was expressed in terms of the dollars of annual sales to a customer newly attracted by each dollar of a moderate amount of advertising. The "death rate" was assumed to be a constant percentage applied to each year's sales.

These three factors were combined in a differential equation ex-

pressing sales for any year in terms of the advertising history up to that time. Solved for values of the three factors provided excellent agreement between actual and theoretically expected sales in several cases, of which an example is shown in Figure 7.

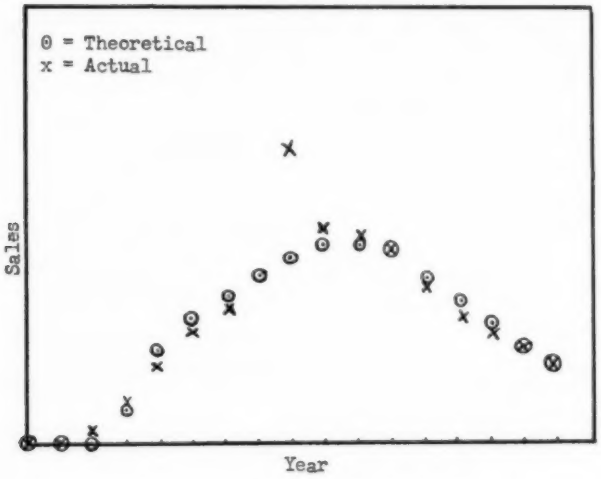


Figure 7

Sales rose rapidly during the early life of the product when advertising expenditures were made at a fairly high rate. At one point it was decided that the product was beginning to go out of style and advertising expenditures were cut back in order to provide money to support more stylish products, and sales dropped. Finally, advertising was stopped altogether and sales fell off at practically a constant percentage rate per year, much like a death rate in a population. The marked difference between actual and expected sales for one year was traced to a type of sales where no influence of advertising was possible.

The fact that practically all of the variations in sales over the life of the product were explained in terms of three factors made it possible to set up procedures for helping the advertising executive budget his available effort after estimating the values of a few of the relevant factors.

These examples represent only simple illustrations of the scope of operations research. It has been useful in other areas such as the evaluation and control of sales promotion programs, the design of optimum distribution systems, and financial planning requiring the evaluation of alternative demands for capital expenditures. The ultimate goal of a company's operations research team--a very distant goal in most cases--is to provide a quantitative model of the operations of the entire enterprise. The framework upon which such a model would be built--and within which the important relationships would be established--may look much like what is often called the du Pont Investment formula[1], illustrated by Figure 8. Here, return on investment has been chosen as the ultimate measure of effectiveness, and the effects of a change in any of the important operating areas of the company can be traced to it.

RELATIONSHIP OF FACTORS AFFECTING RETURN ON INVESTMENT

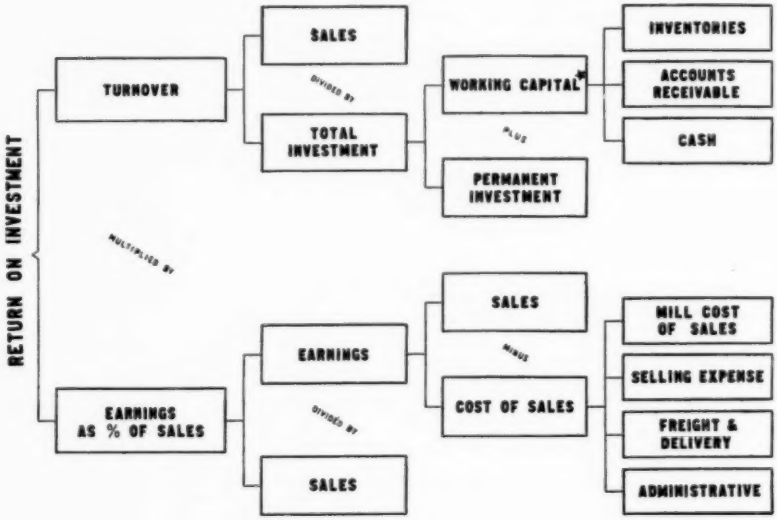


Figure 8

What OR Adds. The question has often been raised as to how the work we have here called operations research differs from some of that which is done within the engineering or research departments of progressive companies, or by management and systems engineers.

At Arthur D. Little, Inc. we have long been engaged in various aspects of development and engineering--particularly those with a technical touch--such as diversification or plant location studies, process flow and materials balance studies, and problems associated with acquisition and merger. And we have compared operations research with these activities in order to try to distill some factors of difference.

One key may lie in the definition of research recently cited by Mr. Mitchell, President of Sylvania--that research is the seeking for knowledge you do not know exists when you start--rather than the application of known engineering principles. It can be distinguished by the kind of people doing it--a research team with a variety of scientific backgrounds. It can be distinguished by the analytical methods required, which are usually non-standard, sometimes rather advanced, and may be drawn from any one of a variety of research fields. A member of a

prominent firm of management engineers, in comparing OR and management engineering in a recent paper, suggests that OR techniques take up where those of industrial engineering leave off.

The emphasis of operations research on the quantitative leads some people to identify it with statistical or accounting analysis. Actual examples indicate, however, that its fundamental objective and contribution is to reveal the basic relationships underlying the operations studied. As in physics and chemistry, theories of underlying mechanisms are sought which are consistent with facts brought out during operations or experimentation, and which can be used to predict the effects of changes in procedure or environment. Like statistics, operations research uses probability theory as a basic tool, but it seems to be ready to employ techniques from many fields, or some which may be specially developed to achieve its objective.

We find, therefore, that successful OR solutions of industrial problems usually result in one or more of the following:

1. New measures of effectiveness.
2. New ways to interpret data.
3. New ways to measure factors not measured before.

where "new" means that they were different from those previously applied to that type of problem.

Developments in Operations Research. The recent formation of the Operations Research Society of America by representatives from industry, government, and universities; the appearance in October, 1952 of the first issue of a Journal for the Society, and the first national meeting in November mark further steps in the growing interest in the application of the methods of scientific research to operating problems of management. The November meeting of the ORSA provided an interesting indication of the scope of present OR techniques and applications. Examples of operations research in production and inventory scheduling, evaluation of costs in machine operations, aircraft maintenance, seaport operations, marketing and agriculture were presented.

Several companies have consciously established operations research groups, some with origin and emphasis in production areas, while others concentrate on marketing problems. Much work similar to operations research has been done under other names. In some companies, development of new products and processes is accompanied by major research effort to measure the impact on sales or operating costs. One organization reported the formation of groups composed of representatives from research, production, and sales divisions to study manufacturing operations and recommend improvements. In the petroleum industry, some research departments have analyzed exploration and production data to provide scientific bases for operating petroleum fields for maximum yield, and for adjusting refinery operations to the best product mix in the light of current price structures and market demands. In some cases, research teams have been asked to provide improved quantitative bases for choosing the best locations for service stations, or for new plants.

The extent to which study by scientists of management problems which arise outside of the research laboratory will expand and be recognized as operations research is not yet clear, but a pattern seems to be emerging. Those organizations which have a long history of substantial research

effort well integrated into the management structure--often evidenced by the creation of subsidiary research corporations--tend to apply first-class scientific study to operating problems up to high management levels. A rough relationship may exist between the degree to which a company's products are research-based and the depth of penetration of research methods into the management operations of the company. The natural growth of research within a company leads to its emergence from the laboratory to deal first with problems of how new products or processes affect sales and costs, then to manufacturing processes and productivity. It is then applied to scheduling of production and inventories, to physical distribution, and later to the evaluation of sales, advertising, and promotional efforts.

When scientific research personnel, using the theoretical and quantitative methods of the basic research, have participated in creating the bases for decisions in all of these management areas for any company, there exists operations research in its most effective form, regardless of what it may be called.

---

(1) How the du Pont Organization Appraises Its Performance. AMA Financial Management Series No. 94.

APPLICATION OF STATISTICAL METHODS IN EVALUATING PERFORMANCE OF ELECTRONIC EQUIPMENT

G. Ronald Herd
Aeronautical Radio, Inc.

It is a well recognized fact that progress in the improvement of equipment reliability has not kept pace with the growth of equipment complexity. In the military services, due to the increased reliance on electronics, inadequate reliability of equipment has become a major problem. Considerable effort is now being devoted to the solution of that problem. One phase of this effort is the present investigation by Aeronautical Radio, Inc. of electron tube performance in military applications. I should like to discuss some aspects of this investigation bearing on the question of the evaluation of equipment reliability.

To evaluate equipments, one must obtain some measure of their effectiveness, and this must ultimately be determined from experience in the field. Only in this way can we be certain of their actual behavior, and only when we know their behavior can we develop equipments which we can be sure will be effective in service. The reliability or measure of effectiveness of equipments is best expressed in terms of probability, i.e., the probability that an equipment will not fail during a period of observation. This can be estimated on the basis of observed data.

Evaluation of equipment reliability is itself a descriptive process and may be accomplished in many ways, such as by observing equipments for a period of time and recording the number of malfunctions or recording the hours-to-failure. The methods used in evaluation are dependent upon the type of equipment, its application, availability of recording time, mathematical ease, and precision desired. Field data in general are observational, i.e., in the nature of results of a survey rather than of an experiment; therefore, the results give overall measures of reliability but usually provide little chance to gain insight into the adequacy of various components or into the effects of factors which contribute to inadequate equipment reliability. To gain insight into the detailed workings of a given equipment type, or to obtain knowledge on the relative importance of contributing factors, it is necessary to supplement survey data with experimental designs.

The carrying out of experimental designs is a difficult problem, because of the large number of variables involved, but the fundamental principles are the same as for any other scientific experiment. The results to be aimed at are a series of numerical answers, representing the dependence of the measures of effectiveness on the pertinent variables. Some of these pertinent variables are the following:

1. Operator
2. Maintenance
3. Complexity of equipment
4. Environment
5. Application

In order to determine the size of the experiment, we must have some idea as to the family of distributions from which our data come, a rough estimate of the parameters, and a value for the degree of accuracy desired. All of these are needed to make reasonable decisions in planning experiments and/or surveys. This in many cases can be obtained from preliminary data derived from past records.

In evaluating equipment reliability, we would like to make certain observations and utilize them to their fullest in making estimates of reliability. Equipment reliability depends on one of the following parameters:

1. Mean time to failure
2. Mean time between failures
3. Mean time to first failure

It is obvious that, in estimating the parameters desired, a continuous variable, such as hours-to-failure, would be preferable to a discrete variable -- i.e., x failures in t hours -- since with the continuous variable, more information is utilized.

In the Military Tube Project of Aeronautical Radio, Inc., our primary effort is to evaluate the reliability of equipments as determined by tube failures only. This is merely one of the many approaches to the overall problem of improving equipment reliability by concentrating on the one type of component which contributes heavily to inadequate reliability.

From field data we have gathered, one may estimate that in 50% of the cases of failure, the repair of equipment includes tube replacements. This statement is based upon information obtained from the three services for various types of equipments and applications. The 50% figure is not the ratio of the number of tubes replaced to the total number of components replaced. It is the ratio of the number of occasions when one or more tubes were replaced to the total number of repair occurrences. Due to the interaction among components within a system, it is quite difficult to isolate that component which actually causes the failure; thus, components such as tubes may be removed when in fact they are not the cause of failure.

Based upon data observed in the field, where the equipment is under normal military operation, we have found that when failed equipments are not replaced, hours-to-failure can be described by the exponential distribution, $f(x) = \frac{1}{\theta} e^{-x/\theta}$.

The θ is the mean time-to-failure, and x is the hours-to-failure. In most cases the data are best described by a gamma distribution, although the exponential distribution (gamma distribution, with $\beta = 1$) fits adequately in all cases and, because of mathematical ease, is preferable. An excellent discussion of the various mathematical models for failure data is given by D. J. Davis.1/

The problem of evaluating equipment reliability is one of estimating the parameter (or parameters) for the distribution of the hours-to-failure of the equipments under study. Thus for the military equipment with which I am familiar, this problem reduces to the estimation of θ, the mean time-to-failure, for the exponential distribution. It

is well known that for the exponential distribution the maximum likelihood estimate for θ after T hours is $\hat{\theta} = \frac{\sum_{i=1}^{r} x_i + (n-r)T}{r}$

where r is the number of ordered failures observed, n is the number of elements under observation and Xi is the hours-to-failure for the i-th element. The ratio $\frac{2r\hat{\theta}}{\theta}$

is distributed as chi-square with 2r degrees of freedom. Epstein and Sobel have discussed this estimate in recent unpublished papers.2/

By utilizing the relationship $\chi^2_{2r}(\alpha/2) = \frac{2r\theta}{\theta}$,

we can obtain an estimate of the number of failures, r, that we would need to observe for any specified value of the ratio $\frac{\hat{\theta}}{\theta}$

The estimate of θ must be obtained from knowledge of the equipment, environment, and other data. If we have an estimate of θ and r, then we can decide upon the length of time, T, we would need to continue the experiment and estimate the number (n) of equipments to place under observation.

Upon completion of the experiment, we can estimate θ and thus we can estimate the probability of failure for any time interval. For example, the expected number of equipments that would complete an x-hour operation successfully without a malfunction would be N (1-p), where p is the probability of failure in the x-hour period and N is the number starting the operation. Now p may be expressed as

$$F(x) = \int_0^x \frac{1}{\hat{\theta}} e^{-x/\hat{\theta}}\, dx \quad ;$$

therefore the expected number is $N[1-F(x)]$.

If we wish to compare the reliability of two different types of equipments, it is necessary to test the difference between the estimated parameters for the exponential distribution. For a discussion of possible tests, I suggest Epstein and Tsao 3/, also Birnbaum 4/; or, for a distribution-free test, I suggest the H- test criterion of Wallis and Kruskal 5/ for testing the difference in mean ranks.

To illustrate the application of the methods just described, I would like to use data obtained in the field surveillance activities of Aeronautical Radio, Inc. We were interested in evaluating equipment reliability as determined by tubes alone for a new equipment containing approximately 60 tubes. It was expected the hours-to-failure for the equipments which failed due to tubes would be exponentially distributed, since we did not plan to replace failed equipments within our experiment. This expectation was based upon data observed in other field operations concerning different types of equipments.

At the time of our experiment, the equipments were under study by the Navy to determine modifications that would be necessary for their adaptation to Navy aircraft. We wished to obtain, simultaneously with the Navy study, information on the hours-to-failure of these equipments, taking into account only those failures due to tubes. Further, we wished to protect ourselves from underestimating the mean time-to-failure by more than 25% of its value. To achieve this degree of accuracy, it was estimated we would need to have at least 50 equipment failures within our period of observation. The best available

estimate of mean time-to-failure was 200 hours. Therefore we felt our period of observation should be 300 hours of operation. It was also decided that 150 equipments would be used, since we could expect a number of equipments to be lost to our experiment because of the continuation of the normal military operations.

The operational procedure in the field for our experiment was to install the equipment in its carrier and, upon failure, to remove it to a central repair shop. Since the carriers for these equipments were all relatively similar, it was agreed that the equipments could be rotated among carriers without invalidating our results. This did, of course, tend to complicate the task of maintaining an accurate account of the time of operation, and thus required additional effort on the part of our field force. Time was kept by flight time, since the military agreed to maintain the equipment in continuous operation during flight while our experiment was in progress. The period of flight was generally of approximately two hours duration so that the hours-to-failure could be recorded quite accurately.

One hundred and fifty equipments were available, and all were placed under observation. Representatives of our organization were assigned to the central repair shop to maintain proper records of time and repairs made. In this case we were interested in time-to-failure without replacement. Our surveillance of the equipment lasted for approximately five months.

Fourteen equipments were lost to the experiment due to inability to accumulate sufficient time and to various military operational reasons, such as the transfer of equipment to another base of operations or damage to the equipment through accidents, etc. Thus, at the conclusion of the experiment, we had 136 equipments which had either failed due to tubes or had completed 300 hours of operation. Specifically, 43 equipments had completed 300 hours of failure-free operation. The primary reason for the large number of equipments lost to the experiment was the problem of obtaining either 300 hours or life until tube failure. The majority of the 14 equipments lost to the test had accumulated only a portion of the 300 hours of operation at the termination of the experiment and they had not had a tube failure.

Since the initial installation of the equipments to the carriers was made at random, it was felt that those equipments which accumulated 300 hours of operation were representative of the whole group. This reasoning was based upon the fact that failure of some equipments to obtain 300 hours of operation was due to limited use of the carriers in which they were installed, rather than any fault of the equipments. Furthermore, the distribution of these equipments by time-of-operation was fairly uniform over the time interval, 0 to 300 hours. Therefore, the analysis was based upon 136 equipments.

The estimate for θ, $\hat{\theta} = \frac{\sum_{i=1}^{r} x_i + (n-r)T}{r}$,

is 276 hours with a 95% confidence interval of $(220 < \theta < 359)$. Naturally we wished to determine if our data could be described by the exponential distribution. This was checked by a goodness-of-fit test in which we utilized 20-hour time intervals. This fit, using a 5% level of significance, was sufficiently good to warrant acceptance of the hypothesis that the data can be described by the exponential

distribution. Utilizing the estimate, $\hat{\theta} = 276$, in the exponential distribution, we can estimate the probability of failure for any time interval. For example, if we had 100 equipments in 100 carriers and we wished to know the expected number which would complete a 10-hour operation successfully without a tube failure, we would have N[1-F(x)] where

$$F(x) = 1 - e^{-x/\theta},$$

hence the expected number would be $100\left[e^{-10/276}\right] = 96.5$.

Thus we would expect 97 of the 100 equipments to function satisfactorily (with regard to tubes) for the 10-hour period. It should also be noted that the 95% confidence interval is (95.5-97.3).

In approaching the task of obtaining greater equipment reliability, it is very difficult, if not impossible, to build up to the equipment through the individual components. Even though we may know the reliability of the components, we cannot, from this knowledge, determine the reliability of the equipment. This is because of the lack of knowledge of the interactions or the interdependency within the system. In the example just given, it is difficult to separate the effects of tube and socket, because of the interdependence of the components within the system.

It is agreed, however, that the steps in achieving increased equipment reliability include determination of the reliability of the components. For if we consider a particular equipment from the standpoint of design, increased equipment reliability implies improved components; however, the converse does not hold, since improvement of a component may have no effect on equipment reliability whatsoever.

Another problem encountered when evaluating the reliability of a system is the desire of many to use more than one criterion for satisfactory performance. This problem can only be solved by using the system performance as the criterion. You cannot define all component failures except through the system performance. This is because there is no exact definition of what constitutes a failure in the case of the gradual deterioration of a component which causes replacement at an arbitrary level.

So far, I have confined this discussion to the problem of evaluating equipment reliability, and have said nothing about improving the reliability of these equipments. Some comments on this subject seem to be in order.

There are four areas in which we can work toward improving equipment reliability. We can improve: (1) The maintenance system and procedures; (2) the installation and environment; (3) the components; and (4) the design and manufacture. At this time, work in any or all of these areas will, I believe, improve equipment reliability.

The major problem in each of these four areas is now quite well defined. In maintenance, it is one of education and training. In installation and environment, it is a problem of obtaining more detailed knowledge of conditions to which equipments are subjected. In the area of components, it is a problem of obtaining more information on component strong points and weaknesses. And in the area of design, it is a problem of designing the equipment more closely around the

components, by which I mean the existing production type components. In addition, the equipment designer should be alert to the possibility of minimizing equipment complexity, i.e., by reducing the number of components. These are two possible ways of putting more reliability into an equipment while it is still on the drawing board.

There should also be some intensive work on the problem of interdependence within the equipment. Interdependence should be reduced whenever possible, but the primary need is for the manufacturer to determine what the interactions are and to formulate some estimate of their magnitude. Following the manufacture of the equipment, well-designed experiments should be used to determine what effect the various components have on reliability; but, even more important than this, the interaction of components must be carefully ascertained. A basic requirement in every case, of course, is that the manufacturer should thoroughly test his completed product by means of well-designed statistical experiments.

1/ D. J. Davis, Analysis of Some Failure Data, Journal of the American Statistical Association, June, 1952

2/ Epstein, B., and Sobel, M. Some Tests Based on the First r Ordered Observations Drawn from an Exponential Distribution, Wayne University Technical Report No. 1, March 1952

3/ Epstein, B., and Tsao, C. K., Some Two Sample Tests on Ordered Observations from the Exponential Distribution, Wayne University Technical Report No. 3, December 1952.

4/ Birnbaum, A., Some Procedures for Comparing Poisson Processes or Populations, with Application to Life-Testing, unpublished paper

5/ Kruskal and Wallis, Use of Ranks in One-Criterion Variance Analysis, Journal of the American Statistical Association, December 1952

INTRODUCTION TO EDUCATIONAL SESSION AND THE VARIANCE RATIO TEST

Edwin C. Harrington, Jr.
Monsanto Chemical Company

Most of us have come to be statisticians of a sort by a most unusual route. It's as though we jumped from an airplane and landed on the 23rd floor of the Empire State Building. That floor is called Statistical Quality Control. To call all statistics lying outside the accepted realm of S.Q.C. "advanced" is a considerable misnomer, because very much of it, and you will find, a large part of this course, is really basic to the advanced and specialized techniques known as S.Q.C.

In fact, as we apply S.Q.C. to our everyday problems, we should soon find our progress seriously limited if we were to explore no farther than the 23rd floor of this large and versatile structure. In the simple interpretation of Control Charts we run head-on into the binomial theorem on floor 5, and tests to show that cycles are or are not present in a control chart can easily push us up to the 43rd floor.

In these first two sessions, we'll take a brief look at some of the lower floors, specifically at "tests for significance" which are the basis of nearly all the superstructure, including control chart methods. Later, you will take a look upstairs into what is known as experimental statistics.

The idea that variation exists in all products, whether they be ball bearings or research experiments on nuclear fission, is well known. No two balls and no two experiments will ever come out exactly alike. All good quality controllers know how to express this variation in terms of standard deviation, or, more frequently, the three sigma limits on a control chart.

In our manufacturing operations we are, of course, interested in keeping this variation as small as possible. If alternate methods of manufacture present themselves, we might chart both methods and select the more uniform method of manufacture simply by taking the method with the narrower control limits. An alternate method, and one which is quite necessary in experimental work as you will see later in the course, is to compare the uniformity by what is known as the Variance Ratio Test.

We are all familiar with the use and meaning of the term "standard deviation". The word "variance" means nothing more than the standard deviation squared, and as such, is simply another measure of the scatter of points about their average. Just as we would select as the best method of manufacture the one that gave us the smaller standard deviation, so the best method of manufacture will have the smallest overall variance.

But here we run into a very practical problem: if you calculate the variance from 5 or 10 observations, and then recalculate the variance from the next 5 or 10 observations, it is almost certain the two variances will be different even though the process has not changed at all between the two samplings. To select the group with the smaller variance as representative of the better process would be foolish, because you know that both sets of observations were taken from the same

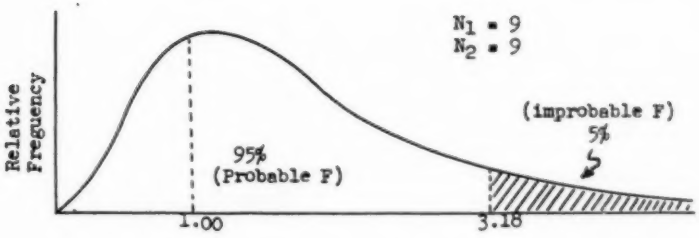
process. This throws a shadow of doubt on the validity of selecting the more uniform process by simply comparing variances, and that doubt is, indeed, justified. We must investigate the problem further.

First of all, let us see what happens to the variance when we take several samples from one process, in which the quality varies strictly at random and follows the normal distribution. For simplicity, let's assume that each sample has the same number of pieces, N. As we noted above, the variance calculated for each of these groups will be different, even though each group is a sampling from the same process. We could plot the variance calculated for each group and get a nice distribution curve representing the distribution of variance of samples of size N taken from our parent population. But this distribution curve would not be of much value since it depends in part upon the true variance of the process. In industrial work, this would be completely impractical because each decision involving variances would require a special curve.

A neat way to make our study independent of the true variance of the parent population is to divide the variance of one sample by the variance of another sample. If these variances are equal, this so-called "Variance Ratio" will, of course, equal 1.00. But since the variances jump around, the variance ratios will scatter above and below 1.00 in a nice distribution. Now it can be shown quite easily that this variance ratio distribution is independent of the variance of the population from which it was derived, and so it is of quite universal application. Because the variance ratio is so important in statistics, it is called the F-ratio, or just F, in honor of R. A. Fisher who was largely responsible for developing the variance ratio test.

The F distribution is independent of the variance of the parent population, but it is dependent upon the number of observations taken in each group or "sample". With only two or three observations in each sample, the variances scatter widely, and so the F distribution is correspondingly broad. With many observations in each sample, the variances are fairly uniform, and so the variance ratio centers closely about the ideal value 1.00. Tables of the F distribution are therefore dependent upon the number of observations from which the variances were calculated . Instead of classifying the F-tables by number of observations, it is more convenient in many of their applications to classify them by "degrees of freedom", which, in the simple case we are discussing, is one less than the number of observations. Thus, if we take samples of 5 from the parent population, and divide one variance by the other there will be 4 degrees of freedom in the numerator and 4 degrees of freedom in the denominator.

These F distributions, for all combinations of degrees of freedom, go from approximately zero at one side to infinity at the other. The F distribution for samples of ten taken from a normal population looks something like this:



Values of F

There are many of these distributions, one for every one of the possible combinations of degrees of freedom in the numerator and denominator. The area under any portion of the curve represents the frequency with which the corresponding range of F will occur. Ordinarily, to save space, F tables list only the value of F which will occur less than 5%, 1%, and 0.1% of the time. Thus, for 9 and 9 degrees of freedom (ten observations in each sample of the ratio), F will exceed 3.18 only about 5% of the time, and it will exceed 5.35 only about 1% of the time if the numerator and denominator are random samples from the same normal population.

We now have a pretty good picture of how quality will vary from a perfectly normal, "in-control" process. So that we are now in a position to select the more uniform process, not overlooking, but taking into consideration the fact that we will find differences even though the samples are taken at random from one single population.

Let us now take a sample from each of two populations with the object of deciding whether the variance (or uniformity) of one population differs from that of the other. If we calculate the variance ratio for the two samples, the F value so calculated will lie somewhere along the F axis of the distribution curve which we have just developed. The problem is to decide whether this value of F would normally have occurred if the two populations were identical or whether this value is so abnormal that we are justified in concluding that the populations are really different.

The logical processes involved in this decision are extremely interesting and fundamental to all of statistical methodology.

We first make the assumption that there is no difference between the variances of the parent populations of the two samples. This assumption of no difference is called the "null hypothesis", which in effect, means that a man is innocent until proven guilty.

Because of the null hypothesis, we may compare our calculated F value with the values of F as they appear in the tables. If the calculated value of F is relatively near 1.00, we have no evidence to assume that the two populations are different, and the null hypothesis, which, assumed there was no difference, is accepted. More specifically, we do

not have sufficient evidence to say that the two populations were, in fact, different.

However, as the calculated F gets farther and farther from 1.00, it becomes less and less probable that the null hypothesis has actually told the truth - it becomes less and less probable that the two populations had the same variance to start with. Somewhere along the line we must conclude that the two populations were, in fact, different. This will occur when F is so large that it is relatively improbable that it could have come from samplings of identical populations.

Unfortunately, there is no sharp demarcation which will say "now it is, now it isn't". We must always deal with the probability that the F we found was large simply by chance alone. By almost universal agreement, "relatively improbable" is defined as occurring by chance not more than one time in 20, or, not more than 5% of the time. This means that if our calculated value of F is as large as, or larger than, the tabled value of F at the 5% level, it is "relatively improbable" that the two variances could have come from identical populations, and the null hypothesis is rejected. That is, we conclude that the parent populations are in fact different, recognizing that one time in twenty this conclusion might be incorrect.

The major difficulty in applying statistical methods, and this includes SQC, lies in the new concepts and the new terminology which are so characteristic of statistical methodology. Once these are mastered, the applications are usually simple and almost obvious.

The following figures represent the tensile strength of a plastic film manufactured by two different methods. Amongst other things, we wished to select the method which would give us the more uniform tensile strength. Although the figures in column P_2 are less consistent than those of P_1, there is considerable doubt that the difference is real, and so we ask ourselves, what is the probability that both sets could be nothing more than two samplings from the same population? (Note that we start with the Null Hypothesis).

	P_1	P_2
	5300	5400
	5600	5100
	5700	5200
	5400	6000
	5400	6100
	5600	5800
	5900	5400
	5500	5800
	5600	5900
	5800	5300
Ave.	5800	5600

We simply calculate the variances for P_1 and P_2 by the usual equation:

$$\text{Variance} = \frac{\sum X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$

and these are 316,000 and 1,289,000 respectively. The variance ratio is calculated by dividing the larger by the smaller, and in this example, this ratio is 3.67.

In both numerator and denominator there were 10 observations, so we have 9 and 9 degrees of freedom. Turning to the tables of variance ratio, we find that on the average, F will not exceed 3.18 five percent of the time, therefore a value of 3.67 is relatively improbable. (The exact probability, calculated by methods not described, is very close to 3%). We conclude therefore, that these are not two samplings from the same population, but rather, that the populations represented by P_1 and P_2 are in fact different, and P_1 is the more uniform. Since we used the 5% level of significance as the criterion of "relatively improbable", we can say further, assuming the data were taken correctly, that there is less than one chance in twenty that this conclusion is false.

* * * * * * *

The foregoing has described the variance ratio, or the so-called F test. I have attempted to show how the F test may be applied as an extension of the control chart method, and indeed it has many applications in problems concerning product uniformity. In plant and laboratory experimentation where very few observations are made the control chart no longer applies, and the variance ratio test is necessary for reliable interpretation of the data.

By means of a few twists and turns, the F test becomes the basis for the analysis of variance and the analysis of variance is the backbone of modern experimentation.

References:

In nearly every text on Industrial or Experimental Statistics there will be found an adequate discussion of the Variance Ratio Test and Tables of the Variance Ratio. To list a few of interest to engineers and research personnel:

Introduction to Statistical Analysis, Dixon and Massey.
Statistical Methods in Research and Production, Davies.
Statistical Methods in Research, P. O. Johnson
Quality Control and Industrial Statistics, A. J. Duncan.

INDUSTRIAL STATISTICS IN PLASTICS EVALUATION

C. H. Adams and E. C. Harrington, Jr.
Monsanto Chemical Co. - Plastics Division

I. INTRODUCTION

It is the intent of this presentation to discuss the scope of product evaluation in the plastics industry and to show how the methods of industrial statistics have contributed to this field. Examples of experimental designs of unusual interest will be given and their statistical aspects described.

II. PLASTICS EVALUATION

Plastics evaluation has as its objective, "Accurate and Complete Definition of Product Behavior in Terms of a Minimum Number of Significant Properties". Obviously, evaluation is not unique to the plastics branch of the chemical industry, but it is of tremendous importance in this comparatively new, rapidly growing business. This is so because plastics materials are sold as packages of useful engineering properties.

Evaluation begins simultaneously with the conception of the idea for a new polymer or even before by showing the need for a certain combination of properties. The chemist justifies his new approach in terms of its projected engineering advantages over existing materials. As the development proceeds, evaluation must be continuously integrated with polymerization studies. Product definition or evaluation serves the same function as do mileage and direction markers for the motorist. It tells the chemist whether he is on the right road and how near is his objective. In addition, adequate evaluation or performance information on the new material serves several other useful purposes: it aids in the forecasting of markets, it insures sound applications and it provides the converter with the necessary fabricating and processing background.

The terminology, "Adequate Evaluation" implies that the product is sufficiently well defined at each step in its development to guide future work whether it be research, production or sales. In the early stages of a research project, this means that a few basic screening tests are called for. As the product moves from the "in glass" phase to semi-commercial scale, more tests are required to bracket processing behavior and predict areas of application. Finally when the production level is reached, evaluation is expanded to define the product as completely as possible. It is a job that is never finished for the longer a material is on the market, the better its processing and use properties must be understood, or it will not maintain a strong competitive position.

The planning of a test program that will provide an adequate evaluation requires sound judgment as to the chemical and physical information necessary. As pointed out above, the amount of testing done is directly related to the level of development of the plastic. Just as an airlines pilot depends on a check chart to insure that his plane is

ready for takeoff, so the chemist should use a check list of properties before he begins his evaluation job. The accompanying check chart for thermoplastics is a helpful tool in choosing the most definitive tests at each stage of product development (see Figure 1). Industrial statistics take over once the tests have been decided upon to provide the most effective, most efficient evaluation possible.

III. THE ROLE OF STATISTICS IN PLASTICS EVALUATION

One of the most frequently expressed reasons for a skimpy inadequate evaluation is that the time required for a good job is not available. Since time is money in any industrial operation, this is tantamount to saying that the right type of evaluation cannot be afforded. The question of whether one can risk the inadequate job and, therefore, whether it is justified at all provides a subject about which a separate paper could be written. The point of importance here is that statistics offering as it does economy of experimentation, even in a complete factorial design, can transform the skimpy inadequate evaluation into an effective adequate one at a lower cost in the long run. A further advantage of the industrial statistical approach is that it forces planned, rather than shotgun or intuitive experimentation. The statistical planning augments planning based on chemical and physical principles. The economy and efficiency of the statistical method is further demonstrated by the technique of partial replication which: (1) reduces the experimentation called for in the complete factorial design, and (2) frequently provides the necessary information before all of the tests in the experimental design have been completed. Industrial statistics enable one to operate at a known reliability level. It defines the calculated risk one is willing to accept. Classical experimentation often overlooks interaction among variables, experimentation based on industrial statistics does not--provided partial replication is kept at a minimum. Statistical philosophy introduces the maximum degree of objectivity into the experimentation. It puts it up to the data to show which variables are effective or controlling and the level at which they operate.

IV. WHAT INDUSTRIAL STATISTICS HAVE DONE FOR MONSANTO PLASTICS EVALUATION

Numerous evaluation projects that have benefited from statistical planning have been carried out in our laboratories. The ones to be discussed here were chosen as being representative of the range of problems encountered in this work.

In the field of new test method development, an experiment was carried out that is of considerable interest from the statistical point of view. The work involved plastic compounds which are widely used as insulators for household and industrial wiring. Their suitability for this application is determined by a series of tests prescribed by Underwriters Laboratories. One of these is long time water immersion at 50°C vs. electrical insulation resistance. Because of the time consuming nature of this test, an experiment was set up and carried out to study the feasibility of screening newly developed formulations more rapidly. The specific objective was to devise a test that could be run on easily and rapidly prepared compression molded slabs that would identify the obviously poor compounds in a matter of hours instead of weeks. The experiment that was designed included seven variables, each

at two levels. These variables included those associated with processing conditioning and material differences (see Figure 2). The variables were arranged in groups as shown. To obtain the same degree of precision by the usual classical design, that is, changing only one variable at a time, would have required 32 x 7 - 224 experiments, a number that was prohibitive. Even a full factorial was not feasible so that 2^7 - 128 experiments would have been required. In the final design, each quarter replication was assigned to one or the other of the materials, allowing the seven factors, all their first order interactions and some of their second order interactions to be evaluated in 32 experiments - 64 tests with check duplication. The results of the experiment indicated the important effects of the variables and their relative magnitudes and served as a basis for further work in this fie field.

The next experimental design to be discussed is of itself not unique but it shows that the importance of statistical concepts has been recognized and applied by a Sales Product Development Group and that industrial statistics is no longer confined to research investigations and production quality control. The problem was to define the optimum conditions for the use of one of our resins. The resin is used in combination with woodwaste to make structural panels and molded shapes. In any such combination, many factors combine to control the properties of the finished object. Economic considerations favor the use of the minimum resin content, short molding time and the most effective catalyst type. These were included in the experiment shown in Figure 3. Instead of a full factorial, a 3 x 3 x 3 Latin Square was carried out. This required but nine experiments as contrasted with 27 for the full factorial. Charge, molding temperature, woodwaste type and molding pressure were held constant. The properties chosen to indicate the most effective combination of levels of the variables studied were specific gravity, flexural strength and water absorption. Flexural strength was weighted the most in deciding which catalyst, molding time and resin content gave the best panel. It was found that pressing time had an insignificant effect, that resin content directly affected the strength as did catalyst type.

Management is continually faced with the problem of maintaining a proper inventory level in achieving its goal of attaining uniform production free of seasonal demand fluctuations. The decision as to what production level to maintain is based on economic experience and information on the storage life of the product. It is an important phase of evaluation of a given plastic to learn what its aging characteristics are under accelerated warehousing conditions. The following experiment was set up and carried out to determine the inventory life of a resin. The plan of the experiment is shown in Figure 4. It will be noted that the resin was aged 0, 1, 4 and 16 weeks and that specimens prepared at each of these times were in turn aged 0, 1, 4 and 16 weeks. In addition, the effect of processing conditions and two different notching tools was studied. The latter was introduced to determine the variability of the most critical operation in the test used to measure the quality level. The results indicated significant effects due to processing conditions and testing (notching tools). They further provided a sound basis for estimating storage life of the resin.

In a growing industry, one is faced with the problem of expanding

the fields of application for the products of that industry. Thus, in the plastics field, there is now a growing interest in broadening the range of outdoor applications. This can be most effectively done by systematically investigating the factors which control weathering characteristics. The following hypothetical experiment is presented to show how one can approach this complex problem. The variables stabilizer amount, plasticizer type and amount, coloring agent type and filler type will be included in the design. The constant factors will be resin and processing conditions. The full factorial experiment for such a study is shown in Figure 5. The tests that would be required to follow the aging of the plastic will not be considered here since they would be a direct function of the material studied. A cursory examination of Figure 5 shows that an almost prohibitive number of samples would be required if the factorial design were to be followed. How then to make the proper choice of experiments? The first move would be to pick the plasticizer that is known to give the best performance over the temperature range encountered in use. The second would be to choose the most effective stabilizer. From this point on statistical methods would be applied to further reduce the experimentation. The final design is shown in Figure 6. The number of experiments has been reduced from 72 to 16 by applying evaluation (chemical and physical behavior experience) and statistical principles.

V. SUMMARY

Just as Lord Kelvin said that it's not a science until a given level of behavior can be assigned a number, neither can a material be considered a product until it is evaluated. Evaluation and industrial statistics work as partners in helping the research chemist develop new plastics, in securing sound applications of plastics materials and in broadening their field of application.

FIGURE I

CHECK CHART FOR THERMOPLASTICS

PROPERTY	TEST	MOLDING POWDERS	INDUSTRIAL RESINS
I – BASIC			
1. Mol. Wgt.	ASTM OR OTHER	x	x
2. Viscosity, sol'n		x	
3. Viscosity, melt.		x	
4. etc.			
II – CHEMICAL			
1. pH			x
2. % Solubles		x	x
3. Organoleptic		x	
4. etc.			
III– PROCESSING			
1. Bulk Factor		x	
2. Angle Repose		x	
3. etc.			
IV – FABRICATION			
1. Hot Forming Temp.		x	
2. etc.			
V – MECHANICAL			
1. Tensile Props.		x	
2. Toughness		x	
3. etc.			
VI – ELECTRICAL			
1. Diel. Props.		x	x
2. etc.			
VII. THERMAL			
1. Heat Dist. Temp.		x	
2. Flammability		x	
3. etc.			x
VIII–OPTICAL AND APPEARANCE			
1. Refractive Index		x	
2. Yellowness		x	
3. etc.			
IX – PERMANENCE			
1. Weathering Characteristics		x	x
2. Dimensional Stability		x	
3. etc.			

FIGURE 2

EXPERIMENT TO DETERMINE SOURCES OF VARIABILITY IN THE MEASUREMENT OF THE VOLUME RESISTIVITY OF PLASTIC ELECTRICAL INSULATION

			A_1				A_2			
			B_1		B_2		B_1		B_2	
			C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
D_1	E_1	F_1	G_1			G_1		G_2	G_2	
		F_2		G_2	G_2		G_1			G_1
	E_2	F_1		G_2	G_2		G_1			G_1
		F_2	G_1			G_1		G_2	G_2	
D_2	E_1	F_1		G_1	G_1		G_2			G_2
		F_2	G_2			G_2		G_1	G_1	
	E_2	F_1	G_2			G_2		G_1	G_1	
		F_2		G_1	G_1		G_2			G_2

PROCESSING

A – MILLING TIME
B – MOLDING TIME
C – SLAB THICKNESS

TESTING

D – TEST TEMPERATURE
E – TYPE OF CONDITIONING
F – CONDITIONING TIME

MATERIALS

C_1 – FORMULATION A
G_2 – FORMULATION B

FIGURE 3

RESIN – WOODWASTE EVALUATION EXPERIMENT,

LATIN SQUARE DESIGN

	X	Y	Z
B_1	FLEXURAL STGTH 6919,5272PSI SP.GRAV. 1.02 H_2O ABS. 23.2% C_1 % RESIN	FLEXURAL STGTH 3277,3283PSI SP.GRAV. 0.961 H_2O ABS. 46.5% C_2 % RESIN	FLEXURAL STGTH 1910,2572PSI SPEC.GRAV. 0.906 H_2O ABS. 83.3% C_3 % RESIN
B_2	FLEXURAL STGTH 4385,3886PSI SP.GRAV. 0.957 H_2O ABS. 56.3% C_2 % RESIN	FLEXURAL STGTH 843,897PSI SP.GRAV. 0.992 H_2O ABS. 100.5% C_3 % RESIN	FLEXURAL STGTH 8307,8630PSI SP.GRAV. 0.928 H_2O ABS. 30.3% C_1 % RESIN
B_3	FLEXURAL STGTH 1757,1958PSI SP.GRAV. 0.908 H_2O ABS. 72.4% C_3 % RESIN	FLEXURAL STGTH 7200,7879PSI SP.GRAV. 0.998 H_2O ABS. 30.3% C_1 % RESIN	FLEXURAL STGTH 3434,3517PSI SP.GRAV. 0.991 H_2O ABS. 84.5% C_2 % RESIN

CODE:

B_1, B_2, B_3 – MOLDING TIME, MINUTES

X, Y, Z, – CATALYST TYPE

C, C_2, C_3 – RESIN CONTENT

FIGURE 4

RESIN STORAGE AGING vs. IMPACT STRENGTH, FACTORIAL DESIGN WITH PARTIAL REPLICATION

		P_1				P_2			
		C_1		C_2		C_1		C_2	
		M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
G_1	S_1	0.394*		0.379		0.439		0.430	
	S_2		0.329		0.347		0.407		0.350
	S_3		0.347		0.358		0.397		0.365
	S_4	0.367		0.337		0.377		0.355	
G_2	S_1		0.324		0.403		0.391		0.397
	S_2	0.428		0.458		0.470		0.447	
	S_3	0.427		0.421		0.426		0.440	
	S_4		0.328		0.286		0.405		0.347
G_3	S_1		0.330		0.342		0.348		0.353
	S_2	0.397		0.330		0.452		0.406	
	S_3	0.376		0.385		0.385		0.359	
	S_4		0.282		0.346		0.347		0.311
G_4	S_1	0.377		0.272		0.406		0.396	
	S_2		0.353		0.303		0.379		0.306
	S_3		0.286		0.320		0.386		0.334
	S_4	X		X		X		X	

* DATA IN UNITS, FT.-LBS./INCH NOTCH

CODE:

G_1, G_2, ETC. - RESIN AGING TIMES

S_1, S_2, ETC. - SPECIMEN AGING TIMES

P_1, P_2 - PROCESSING CONDITIONS

C_1, C_2 - FORM IN WHICH SPECIMEN AGED - NOTCHED OR UNNOTCHED.

M_1, M_2 - CUTTERS USED TO NOTCH

FIGURE 5

HYPOTHETICAL WEATHER AGING EXPERIMENT, COMPLETE FACTORIAL DESIGN (ONE EXPOSURE LOCATION)

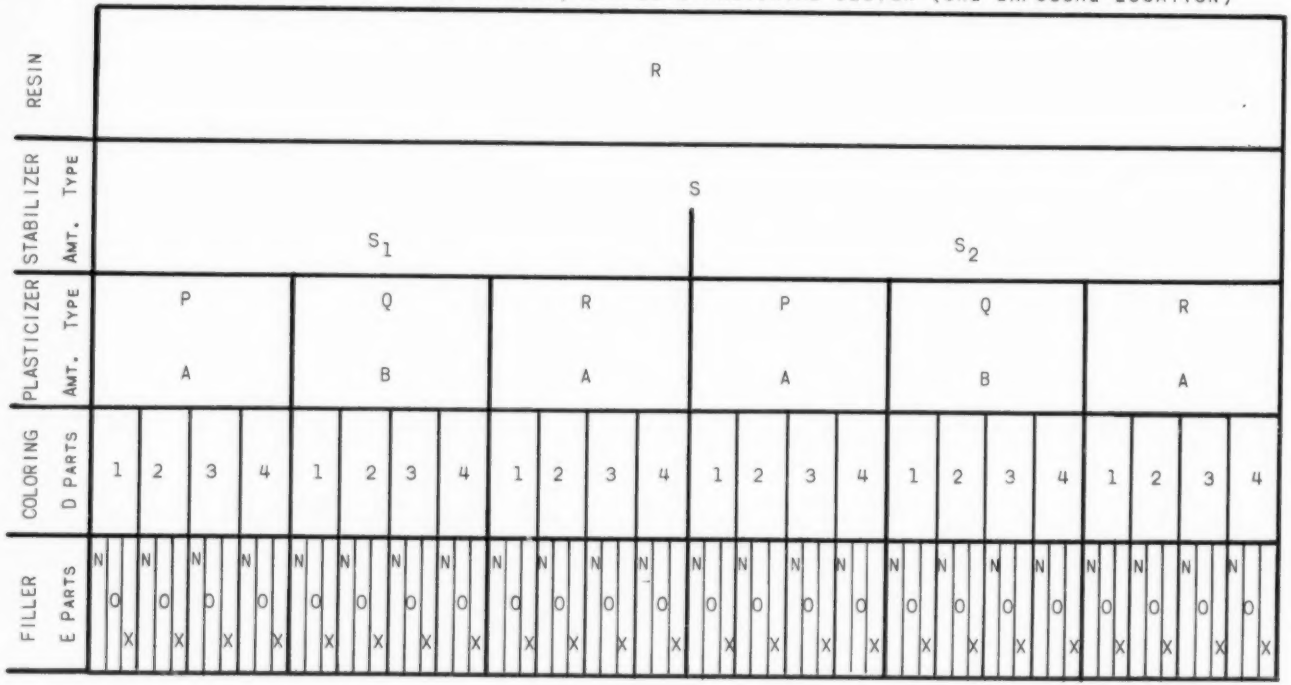


FIGURE 6

HYPOTHETICAL WEATHER AGING EXPERIMENT,
FACTORIAL DESIGN WITH PARTIAL REPLICATION
(ONE EXPOSURE LOCATION)

		A_1		A_2	
		B_1	B_2	B_1	B_2
S_1	F_1	O	X	X	O
	F_2	N	N	N	N
S_2	F_1	X	O	O	X
	F_2	N	N	N	N

CODE:

1,2,3,4 –
COLORANT TYPE

S_1, S_2 –
STABILIZER CONC.

F_1, F_2 –
FILLER AMOUNT

	A_1	A_2
B_1	1	2
B_2	3	4

METHODS OF TRAINING FOR QUALITY CONTROL ENGINEERS AND SUPERVISORS

Ralph E. Wareham
Consultant on Quality Control, Chappaqua, New York

The growth of quality control in industry during the past decade has brought increased need for training key personnel. This training need has been particularly evident in requirements for quality control engineers and supervisors. Even a casual following of newspaper advertisements listing "Positions Open" for quality control engineers and supervisors would show that much training still needs to be done.

When one realizes that ten years ago the total industrial personnel having knowledge of modern quality control methods numbered only a few hundred persons, we can gain an impression of the extensive training that has been done. This training has been carried on in many different ways and through diverse channels. It is not the purpose of this paper to review the training methods used in earlier years, but rather to discuss training methods now available.

Basically, training of a quality control engineer or supervisor involves two basic phases:

1. The educational process of learning to use modern quality control methods, particularly statistical quality control and acceptance sampling; and
2. The practical experiences of applying this knowledge to control of quality of industrial processes and products.

In general, the first step of educational training requires only a limited time, while the second step of practical experience requires a much longer period.

Most of the early quality control engineers and supervisors acquired their training by trial-and-error, in working with factory problems and in studying particular methods applying to these problems. This required a review of available literature on quality control methods, followed by thoughtful consideration as to whether any of these methods would fit the problem at hand. Next a plan for the quality control procedure had to be developed in detail, and reviewed with all interested persons. After usual revisions, the plan would be ready for consideration by the operating personnel of the production unit involved. Acceptance or rejection of the plan was up to the operating personnel. If the plan was accepted, such acceptance was usually qualified and limited; on the other hand, a rejection by operating personnel would mean the end of the proposed plan.

It was on the basis of such trial-and-error methods that the use of average and range charts, per cent defective charts, number of defects charts, process capability studies, cumulative tolerance concepts, and process sampling plans were developed. Scientific applications suggested generalization of quality control methods and techniques. Thus the literature of techniques and of specific applications grew. A good example of such development is found in process capability studies, the techniques for which have been developed within the last few years with new applications reported continuously.

Now the picture regarding the above phases of training quality control engineers and supervisors is completely changed. In the first place, there is an extensive store of literature covering not only quality control principles, but also applications to specific problems in important industries. The recent A.S.Q.C. Textile Division report on quality control applications and methods in textile operations is an excellent illustration of this development. Furthermore, with public quality control courses and organized quality control departments in leading industrial companies, the opportunities for well-rounded training are extensive. What then are the key steps in securing this well-rounded training?

First, the quality control engineer or supervisor must be adept with the modern tools of quality control, and he must also understand the fundamental relationship of these tools to the industrial processes and products of his plant. For example, the quality control engineer in a chemical plant must understand the chemical processing operations as well as quality control techniques. This requires knowledge of two different types.

Usually a number of individuals may have knowledge of one type--either of quality control or of plant processes. What is needed is the combined knowledge of both phases--and this requires training toward one side or the other for available personnel. For example, a recent college graduate may have taken educational courses in statistical quality control but have little experience in industry or with the industrial processes of his plant; on the other hand, a number of individuals in the plant may have adequate knowledge of plant processes but have no knowledge of quality control methods.

In this situation, the question is frequently asked: "Is it better to give a prospect training in the production process or in the quality control methods?" The answer to this question, of course, depends on the aptitudes of the individuals involved. However, other things being equal, experience indicates that the best course is to train the plant men in quality control methods.

Among the reasons for this statement is the fact that qualified men who understand the process can frequently master quality control techniques with relative ease. Furthermore, their background of plant problems, their familiarity with plant personnel, their maturity and experience, and their understanding of company objectives enable them to effect practical applications promptly once the methods are learned. This usually means accomplishment of the objective.

Many educational opportunities are now open to plant personnel seeking training in quality control techniques; principal among these are:

1. Evening courses sponsored by extension divisions of colleges and universities;
2. In-plant quality control courses (either during working hours or on an after-hours' basis);
3. Intensive quality control courses on a full-time basis for eight or ten days, held by colleges, by special groups, or by quality control consulting firms.

Each of these methods of training has certain advantages, but the practical side is that good quality control engineers and supervisors have been trained in all of them.

Where plants are located in or near large metropolitan centers, evening classes are usually available. On the other hand, plants located in smaller communities may be directly dependent on in-plant courses (using outside instructors) or on intensive training courses at distant centers. The particular situation of the individual plant must govern its selection of type of training.

So far, our consideration has been directed to the educational side of quality control training. Once the necessary formal training has been accomplished, the prospective quality control engineer can begin to use his training in practical applications. This is a critical and difficult step. It has been the common experience of many that the gulf between solving class problems and solving plant problems is a wide one. In spite of good intentions and diligent efforts, those assigned to jobs as quality control engineers and supervisors have often failed in technical phases of their jobs due to this difficulty.

The best guidance in carrying class teaching into effective practical application comes from one who has had the experience before. The prospective quality control engineer or supervisor who can work under an experienced quality control manager or quality control consultant can bridge this difficult period with relative ease. Otherwise, an extended period of trial-and-error may be required.

During the period of training in practical applications of quality control methods, the quality control engineer or supervisor needs to meet as many different quality situations as possible. In some companies, this has been accomplished by rotational assignments of quality control engineers to different departments--so as to provide a picture of the full range of quality problems and solutions.

In his work, a newly-assigned quality control engineer will be applying only a small part of his formal knowledge to early quality control studies. This requires that he keep active his interest and knowledge of other methods by continued study and by review of basic methods and techniques. He must keep abreast of the literature by following articles in Industrial Quality Control Magazine and in industry publications. He must strive to extend his knowledge of advanced quality control methods. He must attend quality control meetings so as to have contact with others in the field.

After instruction in many intensive courses and in college teaching for the last twelve years, I have concluded that each student receiving formal training must supplement the training period with a personal effort to broaden his knowledge and perspective through self-planned study. Key requirements for such personal follow-up are training materials covering a wide range of quality control situations and containing quality control problems for practice. These materials should have been thoroughly studied during the period of original instruction.

The literature of quality control has now become so extensive

that the prospective quality control engineer is frequently at loss to know which text materials to use in his personal study program. Use of too many texts or conflicting ones leads to confusion and division of effort. As possible help to those seeking such training materials, the following "kit" is recommended for prospective quality control engineers and supervisors:

"Statistical Quality Control" by E. L. Grant, McGraw-Hill Book Company, New York, 1952

"Sampling Inspection Tables" by H. F. Dodge and H. G. Romig, John Wiley & Sons, New York, 1944

"A.S.T.M. Manual on Quality Control of Materials" by American Society for Testing Materials, Philadelphia, 1951

"MIL-STD-105A--Sampling Inspection by Attributes", Superintendent of Documents, Washington, 1950

"Statistical Quality Control" Article in December 1949 Issue of Fortune Magazine

"Industrial Quality Control Magazine" published by A.S.Q.C.

The above list has been purposely kept short. Later on, the quality control engineer or supervisor could add other valuable texts to his working collection.

The three principal traits needed by prospective quality control engineers and supervisors are a good personality, a spirit of inquiry, and adeptness in simple mathematics. The quality control engineer needs a personality that will help him in his contacts with plant operating personnel; he must "sell" quality control ideas every day. He must have a spirit of inquiry so as to seek out reasons behind quality problems in his plant, and so as to assist in finding solutions. Finally, in our Army Ordnance quality control training schools at Frankford Arsenal, we have found that skill in shop mathematics is a "must" if the quality control engineer is to master statistical quality control techniques.

In many colleges and universities, programs of instruction in quality control have grown to the point where many outstanding men are going directly from college into quality control work. This trend promises to continue and even accelerate.

The combination of better educational training and increased industrial experience opportunities has brought important progress in this training field during post-war years. Nevertheless, the broad problem of quality control training remains before industry and will continue to be important for many years.

STATISTICAL METHODS AND THE ENGINEER

James H. Davidson
General Electric Company

The efforts of most engineers and research workers seem to have progressed during the past century without a substantial amount of aid from statistical methods. However, there has been an increasing appeal to the statician for help, and this appeal seems to have increased in volume and urgency during the past 15 to 20 years. I emphasize the fact that the appeal has come from the engineering and research professions because I do not believe that the statistician should ever be in the position of having to sell his talents in these fields.

The recognition of variation as a law of nature is becoming more and more apparent in scientific work. Our knowledge of natural phenomena and our ability to devise methods of measurement have long since out-distanced our ability to take measurements or to make them agree. In fact, Heisenberg's principle that motion and position cannot be measured simultaneously with accuracy has shown quite clearly that absolute accuracy is unobtainable. The element of chance in experimentation has become one of its most important attributes. In many experiments a measure of the error or random variation is often more important than measurement of the values themselves.

The normal areas in which statistical methods have been accepted as an important aid to the engineer and research worker are the following:

1. Design of experiments
2. Analysis and interpretation of data
3. Correlation and regression.

As in other fields of business, development and research programs should be thought out in advance, and this involves proper design of experiments. All too often an engineering department in its zeal to be economical will strike off on what seems to be a logical path only to change course on a partial answer. Experimental conditions can seldom be duplicated so that when experimental programs are varied or abandoned before their logical conclusion, a great deal of information is inevitably lost.

In the fields of engineering development and research as distinct from process control, the applications of statistical methods are of a more mathematical nature. Once the engineer is convinced that his measurements are subject to random variation, and that there is virtue in adhering to a planned orderly program for experimentation, the way is open for a substantial increase in efficiency.

Most problems involving the design and analysis of a practical experiment require the combined efforts of the experimenter and the statistician. The techniques for design and for analysis have become so complex that the engineer or researcher just cannot spare the time to study many of them. To expect the engineer to become an expert or even an accomplished applied statistician is like asking him to study law so he can defend his own patents or to visit the universities to do his own recruiting. Although he could never do such tasks well and still be competent in his own field, he must know enough about these activities so that he can cooperate with the people in them to his own advantage.

The engineer in industry needs merely to gain an appreciation of what statistical methods can do for him and a general idea of how they operate, before he can fully use a new member on his team - the statistician.

As fields of specialization increase in number and in concentration the need for cooperation also increases. Not only do observations need interpretation, but variations between observations also need interpretation in terms of physical happenings. The "significant difference" from a statistical viewpoint must be evaluated in terms of engineering precepts or the analysis is pointless. Up to a point one can usually say that the more observations made the greater will be the information gathered. However, an economic balance must be made between manpower and money on the one side and useful information on the other. The size of an experimental design can range anywhere from the factorial design (Table I) on the one end to the so-called Latin square design (Table II) at the other.

TABLE I

FACTORIAL DESIGN--(27 EXPERIMENTS)
VISCOSITY OF SILICONE GUM-RUBBER BATCHES

Each Figure in Square is the Sum of Three Determinations
(One for each jar)

	OPERATORS			
Drums	1	2	3	Total
1	447	432	424	1303
2	465	454	450	1369
3	479	498	511	1488
Total	1391	1384	1385	4160

TABLE II

LATIN SQUARE DESIGN--(9 EXPERIMENTS)
VISCOSITY OF SILICONE GUM-RUBBER BATCHES

Each Figure in Square Represents an Individual Determination

	OPERATORS			
Drums	1	2	3	Total
1	151_A	150_B	139_C	440
2	159_B	140_C	149_A	448
3	140_C	183_A	140_B	463
Total	450	473	428	1351

Totals: A (jar 1) = 483, B (jar 2) = 449, C (jar 3) = 419

In a factorial design, all combinations of the variable are tried. As an example of this, take the following actual experiment.

One characteristic used to measure the acceptability of silicone gum-rubber batches is viscosity. It was desired to know whether there was a significant variation between drums of the material (that make up a batch), between operators making the tests, or between various types of viscosity measuring jars. If there had been three drums, three operators, and three types of jars, then 27 runs would be required for a complete factorial design. The results of such an experiment can be shown in the manner of Table I, where each figure represents the sum across the three jars.

Table I represents the sum of the three readings made by operator 1 on drum 1; the sum of the three readings made by operator 2 on drum 1; and so on. Sub totals 1391 and 1384 are the totals of the readings made by operators 1 and 2--the difference between totals results entirely from the difference between operators. The reason is that each of the other variables is represented in an equal manner in the total. In the same way the difference between 1303 and 1369 results entirely from the differences between drums 1 and 2.

From this pattern, one is able to compare the effects of drums--one to the other--and the effects of operators, without fear of any biasing or weighting. Any interacting effect between drums and operators can be measured by noting the total variation among all nine values in Table I--subtracting from the total the variation resulting from differences between drums and differences between operators. The result is a measure of the variation among the nine values that cannot be traced alone to drums or operators (nor could it be traced to jars). It is considered to be caused by any relative effect that might occur between operators and drums (that is, a tendency for operator 1 to read high values for drum 2, while operator 2 reads high values for drum 1, etc.) In addition, the random error is caused by inherent variation in the test method and any other unknown or unconsidered variables. It is considered to act alike upon all values. In the example it would be identical to the interacting effect of all three variables together.

If one next wanted to compare the effects of drums and operators without necessarily considering interacting effects, he would use the Latin square design shown in Table II. Here only nine determinations are run. Referring to the Table, subscript A stands for the first type of jar, subscript B for the second, and subscript C for the third. By comparing the total of the three A's with the total of the three B's, one can see the difference is attributable to the differences between jars. This is because the other two variables considered would be operated in a like manner in both cases. The other two variables--drums and operators--would of course be compared in the same way as before. Again, the error term in this case would be the difference between the total variation of all nine values, and the total variation considered for operators, drums, and jars. This error term, however, would include all interactions and they could not be separated.

The two experiments outlined represent extremes in the manner of investigation for this particular problem. Table II represents a minimum of effort to obtain a given amount of information about the variables. Conversely, Table I represents a maximum effort to obtain somewhat more information; which may or may not be essential in a

specific problem. Between these two extreme approaches lies a whole series of possible experimental designs known as fractional replications. While at present relatively little has been published in the journals or the textbooks about these designs, they are being used more and more in our work because the average problem does not need the extreme treatment. It is the very flexibility of the statistical designs for experiments that makes it difficult to decide on the appropriate one for a particular problem. The experimenter must decide how far he wants to go in choosing between the shotgun approach of Table II, where relatively little is already known about the variables and a quick inexpensive answer is required and the rifle approach of Table I, where one has more information about known variables. The solution for each must be hand-tailored to the particular problem and the proper way chosen among the several that are possible.

In the field of statistical methods, as in many others, one must constantly be in touch with the frontiers of new mathematical developments--indicated by current publications and the thinking of leaders in the field. All too often the applied statistician is called upon to solve problems for which the mathematician or the mathematical statistician has not well prepared the way. One of the most widely recognized and important problems of this nature can be categorized under the heading of "life testing."

The problem of life testing appears in almost every phase of engineering development. The question, How long will the product last under normal usage? is indeed a highly important and common one. A great deal of time and effort is spent in devising and executing tests that will, in a short time, simulate the stresses and strains of the normal life of the manufactured product. It's indeed a serious engineering problem to devise a test so that a refrigerator will receive in a thousand hours the same amount of wear and tear it may be expected to receive in 20 years. Therefore, it's appropriate that some care be given the analysis of resulting data.

In many instances, as typified by the case of the refrigerator, only a relatively small number of units can be placed upon test at any one time. After the first failure it's desirable to discontinue the test because months might be spent in waiting for the second failure. How best to interpret the data resulting from such a test was the first problem undertaken by our statistical methods section. The problem was considered to be one of making the best prediction about the life of the production units represented by the sample tested. Since samples are often costly (as in the testing of resistance to radioactivity), the first question asked by the experimenter is, How many units should be put on test?

In Table III are listed the sample sizes necessary to make various predictions about the material represented by the sample, with various corresponding assurances. One can see from the Table that if 14 units are placed on test and the first failure occurs in 1000 hours, then one could--

Predict with an assurance of 50 percent that fewer than five percent of the production represented will have a life less than 1000 hours

Predict with an assurance of 75 percent that fewer than 10 percent of production represented will have a life less than 1000 hours

Predict with an assurance of 95 percent that fewer than 20 percent of production represented will have a life less than 1000 hours.

TABLE III

k Percent \ P in percent	99.9	99	95	75	50
0.1	6977	4652	3026	1401	701
1	689	459	299	139	70
2	343	229	149	69	35
3	227	152	99	46	23
4	170	113	74	34	17
5	135	90	59	28	14
10	66	44	29	14	7
15	43	29	19	9	5
20	31	21	14	7	4
25	25	17	11	5	3
30	20	13	9	4	2
35	16	11	7	4	2
40	14	10	6	3	2
45	12	8	6	3	2
50	10	7	5	2	1

The number of units required in a sample for a Life Test so as to be sure (with an assurance of P percent) that fewer than k percent of future units represented by the sample will fail in a time shorter than the shortest life observed in the sample.

A similar tabulation can easily be prepared based on the second or third or fourth failure, but in each case the prediction made would be based on the information provided by only one failure. The next problem is how one can use all the information when he has a set of samples in which there has occurred a number of failures. The answer to this involves some knowledge of the type of frequency distribution of the failure times.

The establishment of a likely distribution might be done empirically and its validity limited to the type of product investigated. Or, it might happen that an established distribution, like the normal or the log normal, will have wide enough practical applicability that it can be used to approximate a variety of cases.

Like other tools, statistical methods have their limitations. At the present time, one such limitation from the viewpoint of the engineer is illustrated by the optimal design problem. There are many cases in engineering work where several variables are known either empirically or by theoretical considerations. The problem is to use all of these variables to yield an optimum result in the design of the production.

Of course, such tools as multiple correlation and analysis of variances are helpful in such cases, but generally they leave the engineer less than satisfied. A new method, however, shows great promise of solving the problem. It is called Linear Programming developed largely by Abraham Charnes for the field of operations research. To my mind this might well be one of the most significant advances in the field of application of statistical and mathematical methods. The numerous possibilities for savings and efficiencies to be effected in the development engineering and research fields through the use of statistical methods have been hardly tapped.

A PRACTICAL APPROACH TO LOSS REDUCTION IN PRECISION MANUFACTURING

Edward R. Close
Bausch & Lomb Optical Company

Since the times of Leeuwenhoek (1654) the manufacture of high precision components has been considered a task requiring a high level of human skill and an inevitably high ratio of losses. A common tolerance for the irregularity of a polished optical surface is $\frac{1}{2}$ wavelength of light. Angular tolerances on prisms are frequently given at $\pm 1'30"$, sometimes as small as $\pm 0'30"$. Tolerances of this order approach the limit of mans ability to measure except under laboratory conditions and were the basis of the belief that in producing precision optics one was forced to choose between high cost hand operations with relatively high yields or low cost machine operations with relatively low yields.

The situation in our Precision Optics Division in 1951 was such that in order to meet the delivery requirements of both civilian and military demands, the production facilities would have to be expanded or the production yields would have to be substantially increased. It was decided that a program to increase yields through defect prevention would achieve the desired results. The purpose of this paper is to discuss the basic concepts of this program, the manner of its execution and the degree to which it has been successful.

The organization and management aspects of the program were discussed by Mr. Carl A. Day (1) in a paper presented at the Manufacturing Conference of the American Management Association in Cleveland, December 3-5, 1952.

The basic concept of the approach was the concept of the hidden plant.

Every plant has a hidden plant engaged in the job of making, finding and repairing defective material. The cost of operating the hidden plant is deducted directly from the profits of the visible plant. Any saving that can be made in the size of the hidden plant is the equivalent of adding manufacturing facilities many times as large as the actual savings made.

Analysis of the size of the hidden plant in Precision Optics showed that it was almost half the visible facility. 41% of the productive labor was involved in waste operation. A detailed breakdown shows how productive labor was spent:

Making good optics	59%
Making scrap optics	3%
Making rejected optics	15%
Repairing rejected optics	13%
Sorting scrap and rejects from good	10%
Total	100%

Of course, such tools as multiple correlation and analysis of variances are helpful in such cases, but generally they leave the engineer less than satisfied. A new method, however, shows great promise of solving the problem. It is called Linear Programming developed largely by Abraham Charnes for the field of operations research. To my mind this might well be one of the most significant advances in the field of application of statistical and mathematical methods. The numerous possibilities for savings and efficiencies to be effected in the development engineering and research fields through the use of statistical methods have been hardly tapped.

A PRACTICAL APPROACH TO LOSS REDUCTION IN PRECISION MANUFACTURING

Edward R. Close
Bausch & Lomb Optical Company

Since the times of Leeuwenhoek (1654) the manufacture of high precision components has been considered a task requiring a high level of human skill and an inevitably high ratio of losses. A common tolerance for the irregularity of a polished optical surface is ½ wavelength of light. Angular tolerances on prisms are frequently given at ±1' 30", sometimes as small as ±0' 30". Tolerances of this order approach the limit of mans ability to measure except under laboratory conditions and were the basis of the belief that in producing precision optics one was forced to choose between high cost hand operations with relatively high yields or low cost machine operations with relatively low yields.

The situation in our Precision Optics Division in 1951 was such that in order to meet the delivery requirements of both civilian and military demands, the production facilities would have to be expanded or the production yields would have to be substantially increased. It was decided that a program to increase yields through defect prevention would achieve the desired results. The purpose of this paper is to discuss the basic concepts of this program, the manner of its execution and the degree to which it has been successful.

The organization and management aspects of the program were discussed by Mr. Carl A. Day (1) in a paper presented at the Manufacturing Conference of the American Management Association in Cleveland, December 3-5, 1952.

The basic concept of the approach was the concept of the hidden plant.

Every plant has a hidden plant engaged in the job of making, finding and repairing defective material. The cost of operating the hidden plant is deducted directly from the profits of the visible plant. Any saving that can be made in the size of the hidden plant is the equivalent of adding manufacturing facilities many times as large as the actual savings made.

Analysis of the size of the hidden plant in Precision Optics showed that it was almost half the visible facility. 41% of the productive labor was involved in waste operation. A detailed breakdown shows how productive labor was spent:

Making good optics	59%
Making scrap optics	3%
Making rejected optics	15%
Repairing rejected optics	13%
Sorting scrap and rejects from good	10%
Total	100%

In order to set up a goal in dollars against which to operate, the size of the hidden plant was expressed as a quantity called the "Gold in the Mine" (2). The existing financial and inspection records showed that the cost of operating the hidden plant in Precision Optics was a conservative $500,000 per year of material and labor alone. This meant that if all defectives could be eliminated a half million dollars per year could be saved. However, there is a balance between the cost of defectives and the cost of preventing them which makes it uneconomic to attempt to mine all the "gold". The 1952 goal was set arbitrarily at mining $250,000 of the "gold" or cutting the cost of defectives in half.

The approach then was to release the hidden plant for productive work by reducing the cost of defectives by one-half. Furthermore, it was determined that the improvements made must be permanent. That when the program was completed, controls would be set up that would never permit the reject level to slip back to where it was before the program started. The emphasis was to be placed on the elimination of the causes of defects.

The first step was to find out where the defects were found. Inspectors report their inspection results on I.B.M. cards. A three-months summary of these cards was made to determine which of seven departments were contributing the greatest percentage of defectives. This summary showed that the spherical grinding and polishing department was first with 37%. The plano grinding and polishing department was next with 27% of the rejects. The remaining 36% of the rejects was divided among the other five departments. This analysis clearly indicated that the place to start work was in the spherical grinding and polishing department.

The next step was to find out what kind of defects predominated in the spherical department. The I.B.M. cards for this department were summarized by reason for rejection with the following tabulated results:

Reason for Rejection	Percent of Total Rejects
Scratches & digs	30%
Chips	27%
Thickness	14%
Surface test	6%
All others	23%

Out of 20 possible reasons for rejection there were only four which are generally non-functional defects, led all the others. Surface test and thickness which are functional were the less serious as to percent of total rejects.

The foregoing analysis showed that scratches and digs were the most important source of rejects in spherical polishing and in order to make a substantial reduction in the rejects in this department their cause would have to be found and eliminated.

Of first importance was establishing a truly objective measure of performance for the department with respect to scratches and digs. A percent defective chart did not fulfill the requirements. Whether a lens was rejected or not depended on two factors: (a) the extent of the

damage, and (b) the amount of damage which could be tolerated on that particular lens. In order to objectively measure the incidence of scratch damage without regard to the tolerance, a scratch and dig index was computed on a sample of 500 lenses per week. The statistic computed from this sample was:

$$\text{Index} = \frac{\text{The total damaged area of the lenses in the sample}}{\text{The sum of the diameters of the lenses in the sample}}$$

The scratch and dig index proved to be a reliable measure of the incidence of damage. Real improvements in the process which eliminate causes of scratches result in a lower index. Likewise a higher index when investigated has always resulted in finding an assignable cause. Chart I shows the progress of this index for a portion of 1952.

In tracking down the causes of scratches a list was made of the opinions of people who were closely associated with the problem as to what the causes of scratches were. This list, in the agreed upon order of importance was as follows:

Take-off (removing lenses from block)
Cleaning (removing residual pitch and wax)
Rough milling
Poor grinding
Dirt (foreign particles in polishing agent)
Rough handling
Surface testing (testing surface against master curve on a test glass)

A number of other reasons were offered but the above list was considered to be the most important and therefore the first to be investigated.

Working under the theory that when there are a large number of different causes for defects they do not all contribute equally to the total, the problem resolved itself to designing experiments to discover the most important causes in the list and doing whatever could be done to eliminate those causes.

To discover whether take-off was an important source of scratches and digs 14 lots of lenses were divided in two groups immediately after polishing. Half of each lot were removed from the shell by a chill take-off process, the other half of each lot were removed from the shell by a heat take-off method. Afterward the lenses from each take-off method were cleaned by the same operator using a single cleaning method. The lenses were then examined and the scratch and dig index computed for each half of each lot.

Chart II shows the index for each half of each lot plotted against each other. The diagonal line on Chart II represents the path these points would follow if there were no difference in index between the two halves of each lot. If one method of take-off caused significantly greater damage than the other there would be a predominance of points on one side of the line. It can be seen from the chart that this was not the case. The plotted points were equally distributed on either side of the line and it could only be concluded that there was no

CHART I

SCRATCH AND DIG INDEX

SPHERICAL POLISHING

Index = $\frac{\text{Total Damage Area}}{\text{Sum of Diameters}} \times 10{,}000$

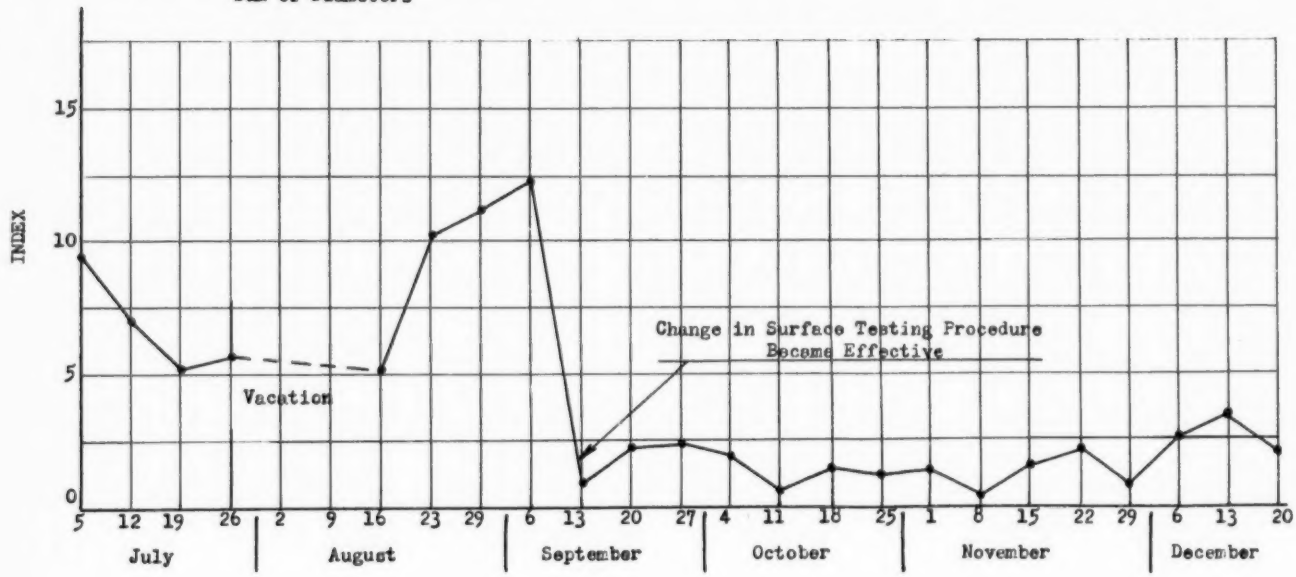
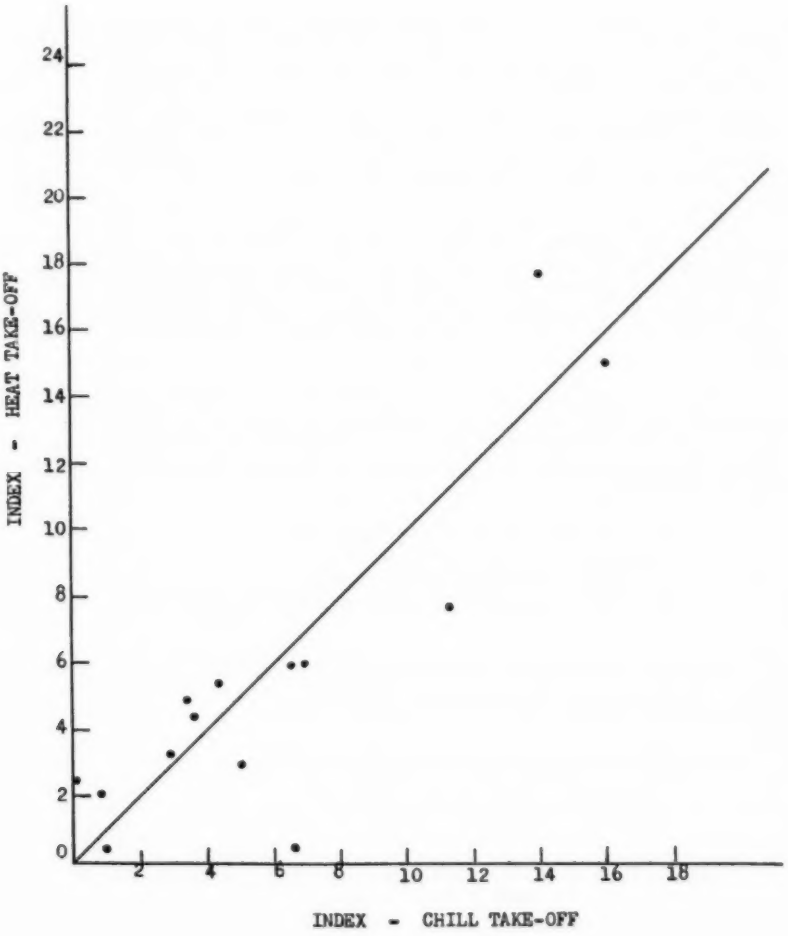


CHART II

Comparison of the Effect of Heat and Chill Take-off as a Cause of Scratches and Digs



significant difference, in the contribution of the two methods of take-off, to the incidence of scratches.

Because these two methods of take-off were so vastly different from each other, it was inconceivable that they could both contribute equally as a cause of scratches. Therefore, neither method was a contributor. From this line of reasoning it was concluded that take-off was not a significant source of scratches and digs. Subsequent experiments proved this line of reasoning to be sound.

By this general method of designing an experiment to determine whether a suspected source of scratches was significant or not a number of suspects from the list were eliminated.

When a comparison of the index between those lenses that were surface tested, versus those lenses not surface tested was made, the true identity of Public Enemy #1 was discovered. The index of the surface tested lenses was 32.6 while the index of the non-surface tested lenses was 2.4. This difference was far too great to attribute to chance. Furthermore, the extent of damage on the surface tested lenses was so extensive it could be concluded that surface testing was the only important source of scratches and digs.

Unfortunately it was not possible to completely eliminate the surface test. However, in making a study of the testing procedure, it was found that by transferring the surface testing operation from an inspector on the bench to a floor inspector in the polishing department, the use of the test glass could be reduced by more than one-half. This change was made and resulted in a sharp drop in the weekly scratch and dig index.

There was some concern at the start of this investigation that if scratches and digs were to be eliminated it could only be done by increasing the direct labor costs. A breakdown of the costs of polishing a typical lens, before the change in inspection procedure was made, are as follows:

To take care of grinding variations	2¢
To polish out the ground surface	7¢
To obtain specified surface test	2¢
To remove scratches (caused by surface testing)	8¢
Total	19¢

Removal of scratches was the greatest single cost in polishing. After the change in inspection procedure was made and the job restudied, the breakdown shows:

To take care of grinding variations	2¢
To polish out ground surface	7¢
To obtain proper surface test	1¢
To remove scratches	4¢
Total	14¢

The reduction in the causes of scratches and digs not only resulted in an improved yield, it also resulted in a reduction of direct labor polishing costs.

The general results of the study of scratches and digs can be summarized as follows:

1. Rejects for scratches and digs were reduced by one-half.
2. The cost of polishing lenses were reduced in general by one-quarter.
3. Changing the inspection for surface test to control inspection on the floor practically eliminated all rejects for poor surface.

The plano polishing department was analyzed to find out what caused the rejects. In terms of dollars the greatest loss was represented by the 42 people who were hand polishing rejected prisms.

Prisms made on the machines are unblocked, cleaned and inspected for performance on a spectrometer or interferometer. Prisms which had astigmatism or power greater than specified were rejected to the hand polishers who corrected the prisms by hand to bring them within tolerance. Many large prisms required over 10 hours of hand polishing labor.

According to optical theory, if a prism is polished with perfectly flat surfaces it should have no effect on the power and astigmatism of the optical system. This is not true in practice. When prisms were polished to within 1/8 wavelength of being perfectly flat their power and astigmatism exceeded the allowable tolerance. This could only be accounted for by some systematic inhomogeneity of the glass.

A straight forward solution would be to improve the annealing process to make the pressings more homogeneous. The glass technicians took a dim view of this. They were convinced that the pressings were annealed by the best process available and any immediate improvement in yield from machine polishing would have to be accomplished at the machine.

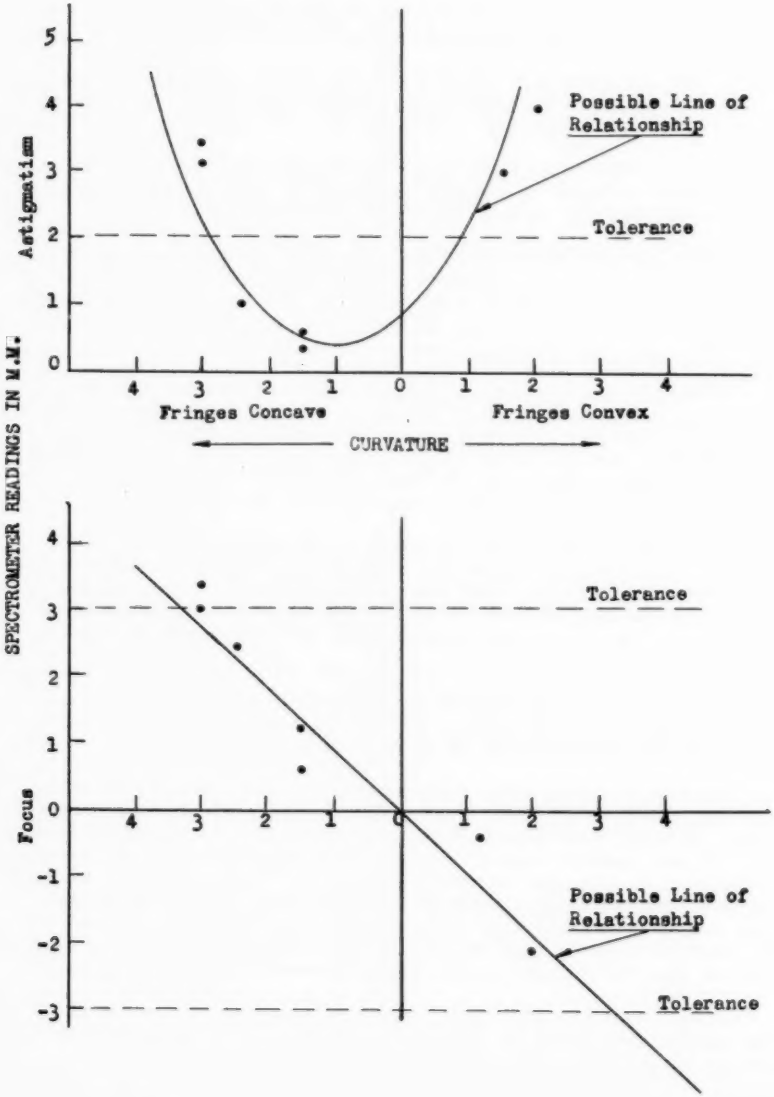
The solution to the problem was reduced to trying to discover what surface, other than plane, to put on by machine polishing which would give the maximum yield of acceptable prisms.

The type of study which eventually provided a solution is illustrated in Chart III. In this specific case seven shells with four prisms per shell were made with the curvature of the reflecting surface from 3 rings (Newton fringes) concave to 2 rings convex. The average spectrometer reading for power and astigmatism for each shell is the ordinate. It can be seen that the lowest readings for astigmatism occur when the surface is 1½ rings concave and, while the performance for power is probably best when the surface is plano, it is not seriously affected when the surface is 1½ rings concave.

As the result of this and many similar studies, a new set of instructions were given to the machine polishers telling them what surfaces to put on at the machine.

CHART III

Relationship of Focus and Astigmatism to
Fifth Side Curvature



The result was a dramatic drop in the amount of hand polishing. Whereas 42 hand polishers were required to correct the rejects from the machine polishing, now only eleven are required for the same number of machine polishers.

The foregoing details are only a small part of the many things that were done to eliminate the causes of rejects and thereby reduce the size of the Precision Optics "hidden plant". The total results of this program can best be judged by the official scoreboards which have been maintained to measure progress.

Chart IV shows the downward trend by quarters of the percentage of rejected lenses in spherical polishing. This percentage was nearly 20% for the year 1951 and reduced to 12% at the end of 1952. This percent of rejections is maintained on a weekly basis for use by the department foreman and is currently running below 10%.

Chart V gives the percent of losses in spherical polishing in terms of dollars. This measure is particularly valuable because it gives a picture of losses that is relatively unaffected by variations in the mix of work. This loss in dollars was 21.4% in 1951 (second quarter) and was reduced to 9.8% at the end of 1952.

Chart VI and VII are the same measures as applied to the plano polishing department. Chart VI shows that the percent of rejects increased during 1951 due to employing inexperienced help needed to fill demands for military production. During 1952 the percent of rejects steadily dropped to 21.8% in the last quarter. Chart VII, the percent loss in dollars, shows why it is necessary to have this record as well as the percent of rejects in pieces. While the percent of rejects decreased steadily during 1952 the percent loss in dollars actually increased due to the difference in mix of work. At the end of 1952, however, it was at the lowest level recorded.

In order to get an overall view, the factory accounting department makes a monthly assay of the remaining yearly "Gold in the Mine". A quarterly summary of this is shown in Chart VIII. At the end of 1952 the goal of $250,000 was reached.

This program, which we consider a practical approach to Quality Control, was launched without publicity or fanfare in September of 1951. In our opinion, it has succeeded to date because of three main factors:

1. The principle of "Defect Prevention" was the prime objective and has been persued tenaciously.

2. The program has enjoyed enthusiastic cooperation from the operating level within the Precision Optics Division and from the flanking staff divisions of the Company. Everyone having anything to do with the specification, manufacture or inspection of precision optics has been a member of the "team" to prevent defects.

3. From the start it has had the understanding and support of

CHART VI

PERCENT REJECTS

-PLANO DEPARTMENT-

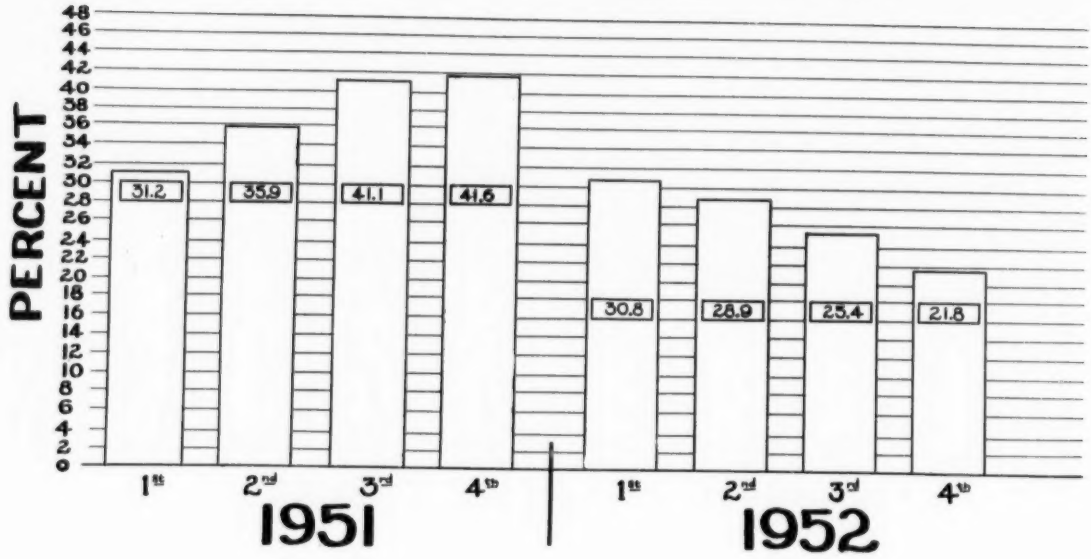


CHART V

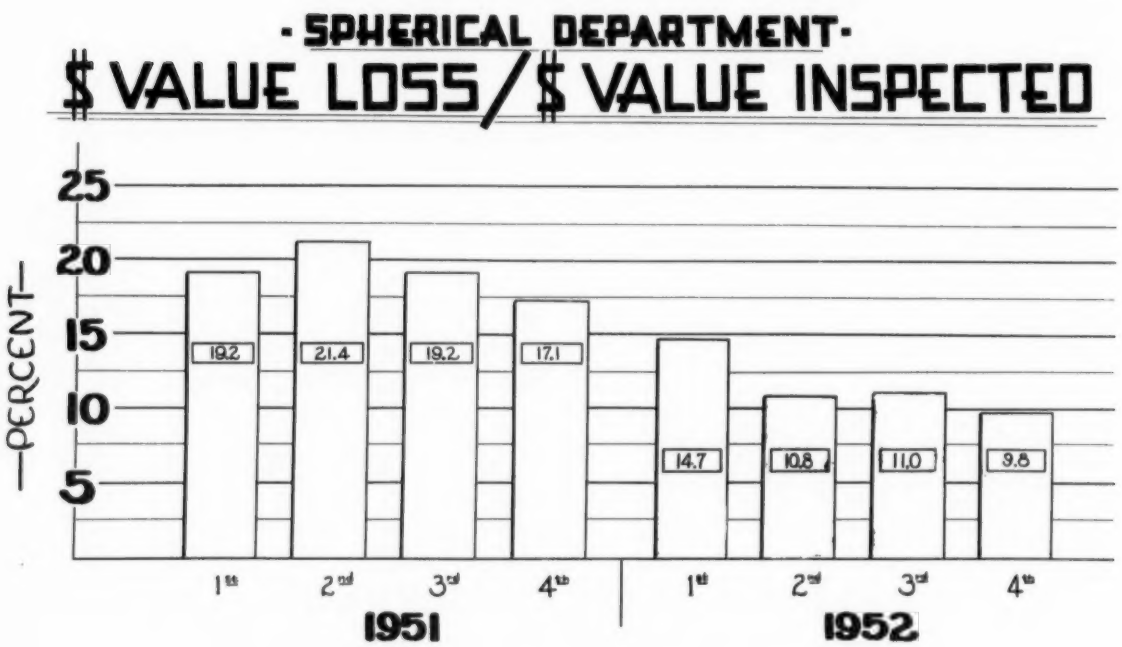

· SPHERICAL DEPARTMENT ·
$ VALUE LOSS / $ VALUE INSPECTED
—PERCENT—
25
20
15
10
5
19.2
21.4
19.2
17.1
14.7
10.8
11.0
9.8
1st
2nd
3rd
4th
1st
2nd
3rd
4th
1951
1952

CHART IV

PERCENT REJECTS

-SPHERICAL DEPARTMENT-

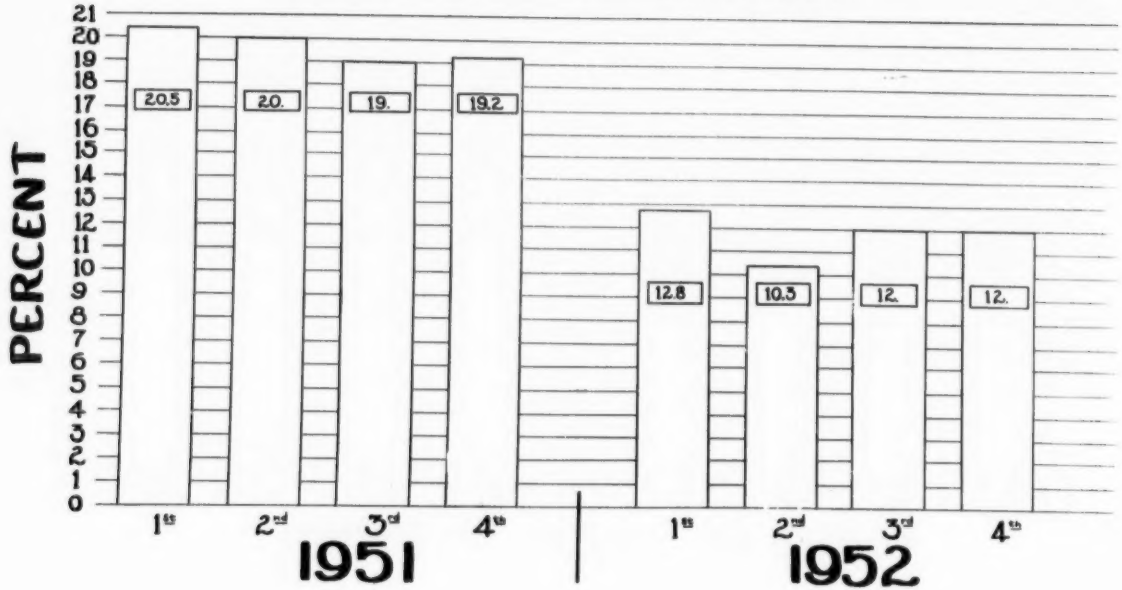


CHART VII

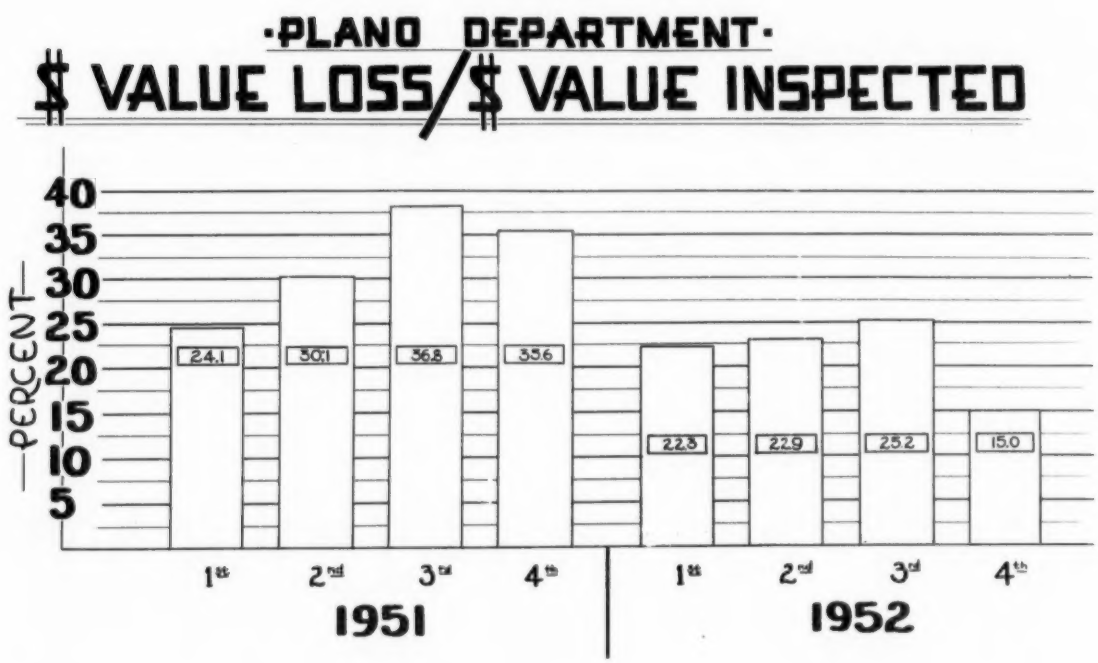
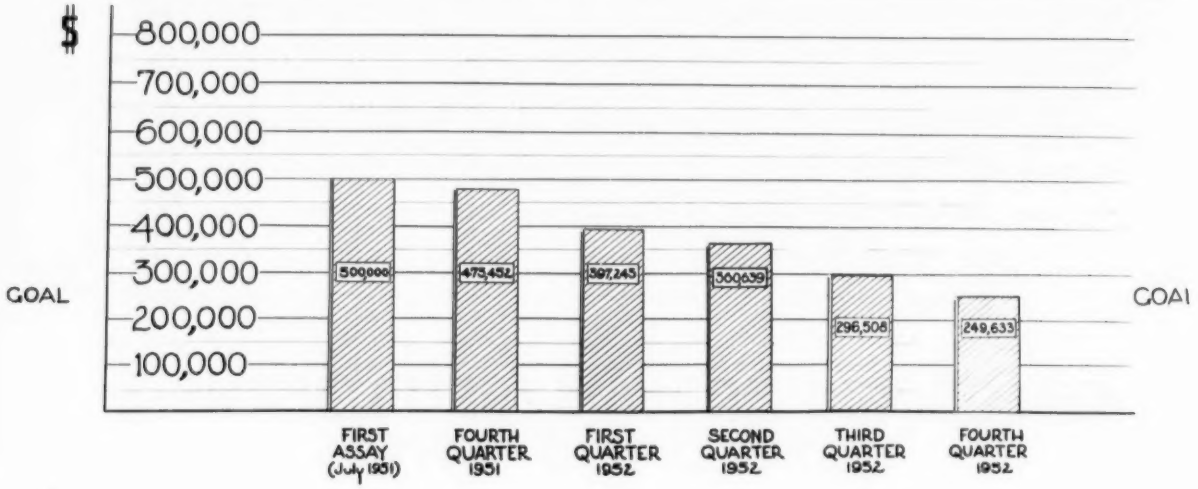

·PLANO DEPARTMENT·
$ VALUE LOSS/$ VALUE INSPECTED
PERCENT
40
35
30
25
20
15
10
5
24.1
30.1
36.8
35.6
22.3
22.9
25.2
15.0
1st
2nd
3rd
4th
1951
1st
2nd
3rd
4th
1952

CHART VIII

HOW MUCH GOLD LEFT IN THE MINE?



the management level of the manufacturing organization.

More work remains to be done. There is still some "Gold in the Mine" that is profitable to extract. Of equal importance is the project of setting up a system of preventive inspection and control such that the gains made can never be lost. This is being accomplished by writing a Quality Control Manual for the Division which details all of the things that operators, supervisors and inspectors are supposed to do to control quality. Nothing is included in the manual until it has been established on the basis of fact to be essential to control. Once it becomes part of the manual, it is subject to a periodic and impartial audit to give assurance that everyone is doing what they are supposed to be doing with respect to quality control.

The reduction of losses, the mining of the gold, and the release of the "hidden plant" have been tangible and profitable results of a modest but successful quality control program. An important but intangible result is the growing awareness on the part of everyone associated with the program that high precision and high losses are not necessarily partners in the business of precision manufacturing. The predictions are that this intangible result will be showing a profit long after the tangible results are forgotten.

References:-

(1) Day, Carl A. "Quality Control: Tool for the Manufacturing Executive. Manufacturing Series, Number 206. "Planning for Efficient Production", American Management Association, New York.

(2) Juran, J. M., Editor - Quality Control Handbook, pp. 37, 337.

SPECIALIZED LIGHTING FOR INSPECTION AND CONTROL

E. A. Linsday
General Electric Company

In many manufacturing processes there are tasks that present difficult seeing jobs. Examples might include the drilling of critically placed holes in metal pieces, or perhaps a machine setting which must be done with a high degree of accuracy. The inspection of leather for surface flaws, or the critical color matching of two or more materials are further illustrations. In many cases, expensive machinery has been purchased so that a job might be completed more quickly and at minimum cost. Labor also represents a goodly portion of the total costs of a product. Therefore, anything which will increase the efficiency of the worker will inevitably be reflected in increased production and lower spoilage costs. Well designed lighting will often perform this function.

The specialized forms of lighting for inspection and control are often called supplementary lighting. The name is particularly fitting since this lighting should always be designed to supplement a good general lighting system. It is poor practice to use supplementary lighting as a substitute for general illumination. A good overall system is obviously needed to supply adequate illumination for the many general seeing tasks in the factory. In addition, general lighting contributes toward providing a comfortable working environment. This is illustrated in Figure 1 and 2. The girls task is the proper forming of projection lamp filaments. This operation necessitates seeing fine details on shiny wire with great accuracy. Figure 1 illustrates a makeshift supplementary lighting system, with the complete absence of general illumination. The seeing problem is complicated by direct and reflected glare from the poor supplementary lighting units, but aside from this fact, the surroundings are gloomy. Each time the girls look up from the work their eyes are forced to adjust from the high brightness of the work to the darkness of their surroundings. This adjustment occurs many times

Figure 1

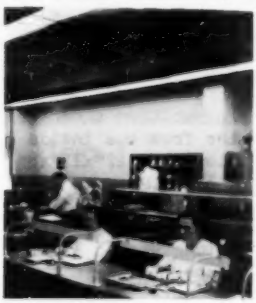


Figure 2

each day, and may result in visual fatigue. Compare Figure 1 with Figure 2 where the same task is being performed under a good system of general illumination, and with the addition of well designed supplementary lighting. Also the walls, ceiling, and other interior surfaces when finished in light colors aid greatly in making the visual environment more comfortable. The high contrast between the visual task and the surroundings has been reduced considerably.

Supplementary lighting may be applied in four general ways. The first merely employs additional light, which should come from a well shielded source, with no special beam characteristics required. The second method involves the use of some form of directional light source. A relatively large diffuse light source, uniform in brightness, is the third type. The fourth way includes those highly specialized applications such as black light, polarized light, optical magnification, etc.

Inspection techniques employing additional light are most often useful when the object or detail to be inspected is small in size or has very little contrast with its background. Figure 3 shows the use of a two-lamp, 40-watt, well-louvered fluorescent fixture to provide additional light for examining the soldered joint on a bi-metallic strip. The unit has been mounted in such a way that the increased light is achieved with comfort. This inspector has over 200 footcandles to aid her in her work.

Defects in spring steel are often hard to see because of low contrast. The problem has been solved in this plant by means of a 20 watt fluorescent lamp in a trough type reflector. (See Figure 4) The lamp provides adequate footcandles for the operator to inspect the steel.

Another unusual application of light for inspection is shown in Figure 5. This tank enables the inspectors to easily and quickly determine the location of any leaks in meter assemblies. The assembly is connected to an air line at 80 pounds pressure and lowered by a hand crane until submerged in the tank. Location of any leaks is clearly revealed by bubbles which are lighted by a 100 watt inside frosted lamp in a vapor-tight fixture. One lamp is installed in each corner of the tank, and the tank is painted white on the inside for improved light distribution.

Figure 3

In another plant supplementary light is used to improve the seeing of the "benching" operation shown in Figure 6. It consists of cleaning the weld spatter from the inside of metal cylinders. A 40-watt filament lamp, mounted in a 12"x12"x8" box with an opal glass cover provides the source of light.

A lamp attached to a long handle is often used in getting needed light into normally dark areas. Figure 7 shows the use of such a device to check the interior of a bomb casing. The lamp is a standard 60-watt filament lamp in a wire guard mounted on a long piece of conduit. It is shielded in such a way that direct light from the lamp will not strike the operator in the eyes.

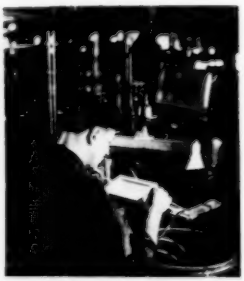


Figure 4

The second type of supplementary lighting makes use of a directional beam of light. This technique is especially useful for detecting surface flaws, irregularities, pit marks, scratches

Figure 5

Figure 6

and cracks in non-specular materials. When the directional beam of light strikes the surface to be inspected at an oblique angle, the defects show up as highlights and shadows. Sometimes specular materials may be inspected in this way. The surface to be inspected will appear dark to the operator, except for the imperfections which will be bright, since they will reflect the light. Care should be taken to position the source of light in such a way that the reflected beam does not become bothersome to the operator or the other personnel.

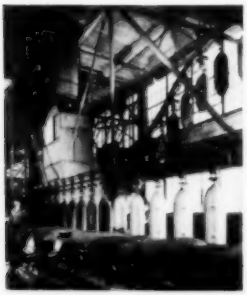


Figure 7

Figure 8

In order to produce the desired directional beam of light for this type of supplementary lighting, narrow beam reflectors, lenses, or spot lamps are generally used.

In a paper company plant five all glass sealed beam automobile fog lamps are used to show surface defects on coated paper sheets. (See Figure 8) Light is directed along the surface of the paper and the operator is not bothered by glare from the lamps as fog lamps have a very sharp cutoff.

The same technique is often used for the inspection of sheet metal as shown in Figure 9. Eight 150-watt reflector spot lamps on 6-inch centers light these etched zinc plates for critical inspection in a lithography shop. The lamps are adjustable so that the best grazing angle can be set. Defects show up clearly and quickly.

Figure 9

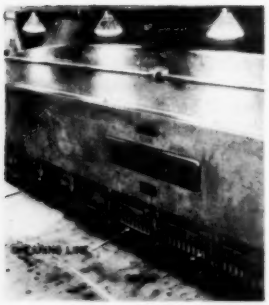


Figure 10

PAR lamps inherently produce more concentrated beams than the reflector type lamps. In Figure 10 they are used to project a shadow or guide line on material in a shearing machine. The 150-watt PAR-38 lamps are mounted in such a position that light is directed through a slot along the edge of the shear blade. A shadow is formed at the point where the blade will shear the metal. The operator can then place the material in the proper position for cutting.

Edge lighting is another good inspection technique which makes use of the directional quality of light. Figure 11 shows a sheet of plastic which is illuminated through its edge by light coming from a slot in the box below the operator. The light is provided by an 8-watt fluorescent lamp behind the slot. When viewed against a dark background such as black velvet, surface scratches as well as internal bubbles in the plastic can be easily detected. The technique is also employed with some types of glass.

The third technique used in supplementary lighting makes use of a diffuse light source, generally large in area and uniform in brightness for illuminating the material to be inspected. One very common use of this technique is in the photographic and printing industry. For example, negatives are transilluminated to make the details easier to see. In Figure 12 this technique provides a good method for inspecting color-separation negatives in the printing industry.

Seams in any light fabric, such as those in a parachute are more readily inspected when transilluminated. (See Figure 13) The design of the source of light is much the same as in Figure 12. It consists of either filament or fluorescent lamps in a properly constructed enclosure which has been painted white on the inside and covered with translucent plastic or glass.

Figure 11

Many food industries use a similar method for careful inspection of their product. Here (Figure 14) the product is checked using the transillumination technique. The luminous panel behind the bottles quickly reveals any foreign material in the bottles as well as partially filled bottles.

Another unusual application of transillumination is shown in Figure 15 where maraschino cherries are being inspected for pits before bottling. A green

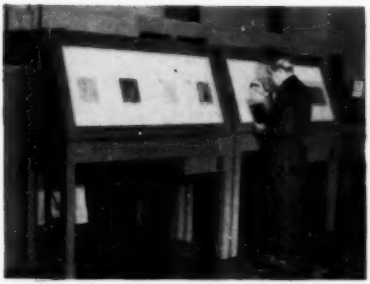


Figure 12

Figure 13

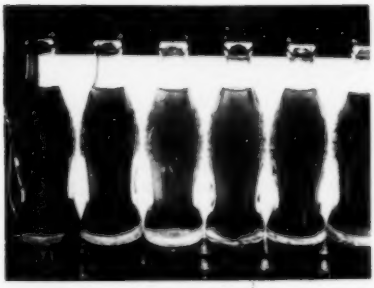


Figure 14

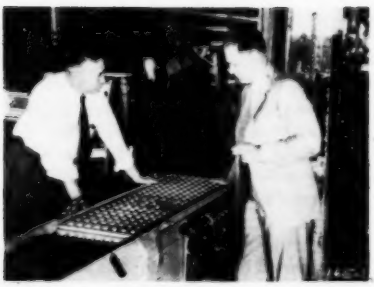


Figure 15

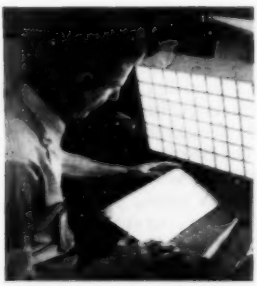


Figure 16

fluorescent lamp is used behind translucent glass under the conveyor. The pits, of course, transmit no light and are seen in silhouette.

Controlled reflections provide us with still another inspection technique. Figure 16 shows the same type of low-brightness diffused source which we used for transillumination being put to a different use. In this case a criss-cross pattern has been painted on the translucent plastic panel. The reflections of this source from a specular material reveal the presence of dents in the material. Irregularities in the reflected lines indicate these surface defects. The material must be moved to be sure that the whole surface is covered with the pattern. This particular unit contains three 15-watt fluorescent lamps.

The finish on automobile bodies may be inspected for scratches and dents by means of controlled reflections. (See Figure 17) This installation consists of a tunnel formed by banks of two-lamp industrial fluorescent units. As the car is driven slowly through the tunnel, inspectors watch for distortions in the reflected pattern of the light sources.

Some types of inspection require techniques mentioned in the fourth catagory.

Figure 17

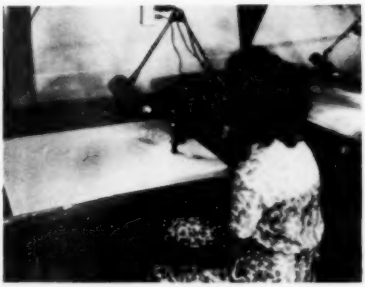


Figure 18

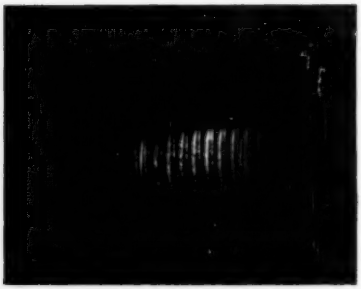


Figure 19

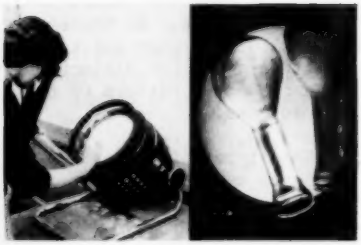


Figure 20

Black-light, the popular name for the near ultraviolet radiation which causes many materials to fluoresce can be used in a variety of ways. Figure 18 shows its use in a sewing operation. The pattern is stenciled with chalk which fluoresces under black light, but which is invisible in ordinary light. A 15-watt fluorescent type black light lamp is used to provide the ultraviolet energy.

One of the chief uses of black light is in the metal industries. For the inspection of threaded shafts, they are first magnetized, then sprayed with a suspension of fluorescent ferromagnetic particles in oil, and then viewed under black light. (See Figure 19) Cracks cause leakage of magnetic Flux, which induces a concentration of the particles at the crack. This concentration is revealed by black light. Non-magnetic materials may also be inspected for surface flaws with a slightly different technique.

Figure 20 shows the inspection of glass bulbs with polarized light. Internal strains in glass appear as streaks with color fringes. The polarizing screens in this particular instrument are located in the eye piece and directly in front of the light source - which is a filament lamp in this case. This type of inspection can also be used on radio tubes, lenses, transparent plastics, etc.

(Figure 21). Optical magnification is provided by this inspection unit which uses a circline fluorescent lamp as its light source. Here the operator can see very fine detail.

Figure 21

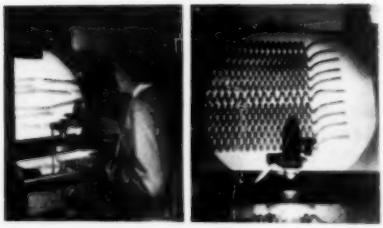


Figure 22

Another example of optical magnification is shown in Figure 22. Lamp filaments on the table are lighted from below and their silhouette is magnified and projected on a screen for critical inspection. Slight defects can be readily identified.

In four-color printing it is difficult to inspect the various color proofs. The yellow proof especially, has little contrast with the white background and is, therefore, very hard to see. If this proof is viewed under blue light, however, the yellow portions of the proof will absorb most of the blue light and appear dark, while the white portions of the proof will reflect most of the blue light and will appear bright. The contrast between the proof and its background is thus increased and much greater detail can be seen. Figure 23 shows a color-proof examining booth which employs this principle. The ceiling is composed of banks of various colored fluorescent lamps. Green light is used to check the magenta (red) proofs, red light for the cyan (blue) proofs, and blue light, as mentioned, for the yellow proofs.

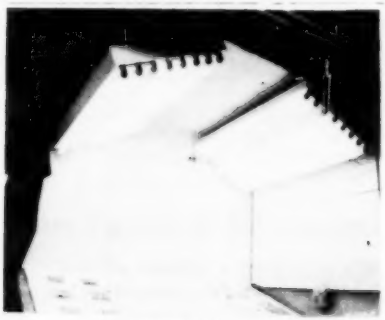


Figure 23

Critical color matching, requires the use of two light sources of markedly different spectral distributions. If two paint samples, for example, will match under light from a filament lamp and also under light from a daylight fluorescent,then they will match under any light source. For critical color matching the source of light shown in Figure 24 can be used. This source combines daylight and blue fluorescent lamps plus silvered-bowl filament lamps. By switching from the fluorescent lamps to the filament lamps satisfactory matches can be assured.

The above illustrations serve to give some idea of the many places where supplementary lighting for inspection of difficult seeing tasks help insure easy accurate seeing. In general the technique for a certain job is developed after careful study of the problem. The

Figure 24

application of fundamental principles in the control, color quality, and quantity of illumination contribute a great deal toward the elimination of the seeing "bottlenecks" which in many cases are the most troublesome features of manufacturing processes. Accurate comfortable seeing means better inspection and easier quality control. When a worker Sees Better he Works Better.

(No representation is made concerning the patent situation involving any of the above described inspection techniques, and any prospective user should make his own investigation thereof).

PRETESTING PACKAGED PRODUCTS FOR SAFE TRANSIT- A VITAL PHASE OF QUALITY CONTROL

R. F. Bisbee
Westinghouse Electric Corporation

On the premise that "all engineering, manufacturing and quality control efforts are in vain if the product reaches its destination in a damaged condition," then it is obvious that pretesting of the finished packaged product is a vital phase of Quality Control that has been long overlooked.

Quality Control techniques go to great lengths on the control of raw material, material in process and the finished product, but usually that is where it ends.

The purpose of this paper is to clearly set forth and prove that it is just as important to control the packaged product by predetermining if it will stand normal handling and transportation, as it is to control any other phase of the product through manufacture, in that subsequent inspection after pre-shipment tests reveals defects that cannot be found in any other way.

The benefits of pre-testing or Quality Control on the finished packaged product can be divided into four parts.

One - A Tool For Better Quality Control

To control and maintain a uniform quality of the product by checking and testing a sample of each product produced daily after the Safe Transit tests have been conducted.

Two - A Tool for Reducing Product Losses in Transit

To assure the product will stand normal handling and transportation before it is shipped. By doing so, it guarantees the product will reach its final destination in good condition.

Three - A Tool for Reducing the Costs

By having a definite standard for pre-shipping tests, it becomes a positive tool in reducing the costs to the "last drop," and still be safe for shipment.

Four - A Sales Tool

It is obvious by the benefits gained listed above that this program is a definite sales advantage in showing the customer that the shipper is doing everything possible to save him trouble and expense by reducing damaged goods.

BACKGROUND OF NATIONAL SAFE TRANSIT TESTING

Before going into detail on the benefits derived by using the pre-testing packaged product program, it will be necessary to give you a brief outline of its background and some of its details in order to give you a complete picture. To assist in this, later in the program I will show you a 30 minute colored sound film which will give you all the details of the testing procedures, etc.

What It Is

The National Safe Transit Program is a cooperative industry plan which was organized to reduce shipping damage losses, which has also resulted in other benefits which will be outlined in this paper. To achieve this end, a number of associations representing manufacturers of major appliances and allied metal products, transportation agencies and container manufacturers, have pledged their full support.

The Problem - Shipping Damage

Losses due to damage sustained in the shipments of goods comprise a heavy and wasteful burden on the American economic system.

It is estimated the total losses of all carriers and refused claims are about 250,000,000. A vast effort, as well as money, is required to report, investigate and adjudicate such claims.

Costly legal proceedings must often be invoked. The harmonious relations between the manufacturer, his distributors and dealers upon which buying confidence is based are often disturbed and the resulting loss of goodwill injures everyone.

Early Research Work

In 1936 we made a check with industry and leading laboratories, and found that there was no standard for a pre-shipment test of a packaged product.

We found that many attempts had been made by various groups to do something about lowering losses, but because they had no actual approach to the solution, no tangible benefits resulted from these campaigns.

In that same year we undertook to establish and correlate a standard pre-shipping test. To make a long story short, by 1941, working with other manufacturers, we had developed a pre-shipping test and had proven its correlation with actual handling and shipping. As proof of this, our damaged product losses shipped had been reduced from 3% in 1937 to about 1% in 1941.

National Safe Transit Committee Organized

After the war our work was continued, and as of August 9, 1948 a committee was formed known as the National Safe Transit Committee sponsored by the Porcelain Enamel Institute of Washington, D. C. Details of this organization will be given later on in this paper.

Therefore, this pre-testing procedure became the first national standard in history on major appliances and allied metal products.

The Search for a New Approach

The committee was fully aware from the beginning of the magnitude of the task it had been assigned. Many attempts over many years to find a solution to the transit damage problem had been made. The carriers, trade associations, commercial laboratories, research organizations and individual manufacturers had carried on almost innumerable continuous investigations without producing any acknowledged solution of general applicability; nor had all of these efforts succeeded in making any material reduction in the amount of the losses.

A New Approach Needed

The findings and reports of these investigations constitute a comprehensive library of literature on the subject of shipping damage with which the engineers on the committee were familiar.

They knew, however, while this data was valuable for many purposes, they could not provide the tool we needed for our task. To achieve positive, tangible, measurable results we had to find a way to prevent damage before it occurred.

We were not interested in diagnoses that would effect a "cure" after damage had happened. Fortunately, our experience and research work, mentioned above, pointed the way to this approach.

Analysis of Causes of Shipping Damage

The causes underlying shipping failure may be many and complex. Although we speak of "shipping damage," factors other than those involved in actual transportation may bear a heavy responsibility for the failure. Poor packaging, weak product design, carelessness in manufacturing, all these may be contributory causes.

It was obvious that a solution that failed to integrate all these factors could not work, and an engineering approach to the problem, therefore, was necessary. This concept of the application of an engineering analysis of the problem became the keystone of the Safe Transit standard procedure.

Essentiality of Cooperation

The nature of the task to be performed made it obvious that a program to reduce shipping losses had to be a cooperative effort between industry on one hand and the carriers on the other. It was necessary that there be recognition by each that it shared a responsibility for the losses.

Because Westinghouse Electric Corporation was one of the first to standardize on preshipment tests for major appliances and because of our extremely low field transportation losses, I was selected as chairman.

The other men appointed are shown in our Organization Chart. (Attachment #1)

Note: In that this paper is presented primarily to show how pre-testing of the packaged product is a vital phase or tool of Quality Control, I will analyze this phase of the program first. Also because it is my understanding that the main purpose of these papers are to outline all other benefits, I will also outline in brief form the other three benefits derived from adopting this national standard or pre-shipping test that is applicable to any kind of packaged product.

ONE

A TOOL FOR BETTER QUALITY CONTROL

The true quality of a product may appear to meet standards after its completion. Many of our leading quality control organizations in many industries stop quality control at the end of the assembly lines. At this point the true quality control cannot be definite until after it is subjected to Safe Transit pre-shipment tests which simulate actual handling and transportation.

I will attempt to prove this.

With the many variables in material, material substitutions and workmanship that occur each day on any kind of product, many irregularities appear after transportation tests are made that are not visible before.

The Safe Transit pre-shipment test first softens up the product by the vibration test that simulates the resonance incurred enroute by railroads and trucks. After vibration the subsequent longitudinal shock which simulates handling and switching of cars, shows up a weakness that would not be disclosed in any other way until after many thousands are made and reported back from the customer.

Examples:

Westinghouse Electric Corp., Mansfield Works

The other day a welding buck slipped in the manufacture of our electric porcelain enamel ranges. The packaged range that was tested for that shift by the National Safe Transit pre-shipment procedure disclosed a large enamel chip at the point of stress that developed because of the brace being misplaced in the welding buck. Immediately we checked several others and found the same failure. The trouble was traced back and the welding buck was repaired. As a result no additional products were made that would cause failures. Had it not been for this pre-shipment test, many thousands would have been shipped before we would have received word of this trouble in the field.

Continental Water Heater Co., Los Angeles

K. Eisinga, Vice President -- Points out that National Safe Transit tests conducted at the AGA Laboratories in Los Angeles revealed certain weaknesses that have since been corrected, which improved the product quality.

Day & Night Division, Affiliated Gas Equipment, Inc., Monrovia, Cal.

V. R. Mottinger, Chief Process Engineer -- "...we feel that the National Safe Transit Program is proving of considerable benefit to our company.

The benefits derived by our company are actually in several directions. First and foremost, we are sure when a package is approved by the testing procedure, that we have the best and cheapest package available. Our own plans are to continue to check each of our packaged appliances in an endeavor to get each approved under the program. In some cases our existing package passes with flying colors. However, it has been most interesting to find that there are some inherent weaknesses in our appliances which are showing up in these tests, and which we have failed to recognize through the usual medium of customer complaint. We anticipate producing a better product as a result of the use of this program."

Package Research Laboratory, Rockaway, N.Y.

Earl R. Stivers, Director -- "As an NST-certified laboratory, we have had the occasion to make many National Safe Transit tests. The interesting thing about these tests is that frequently they uncover weaknesses in the product which is being packaged.

We had one article which disintegrated after only 30 seconds on the vibration tester. The manufacturer was so impressed with the possibilities of this test that he stopped production on that particular item and instructed his engineers to test all new products on the vibrator before going into manufacture. There was a second item which we tested that withstood just 15 minutes before falling completely apart. In both of these cases the weaknesses in the designs were quickly uncovered. These are two examples of the effectiveness of this National Safe Transit test, but they are not isolated cases by any means."

TWO

A TOOL FOR REDUCING PRODUCT LOSSES IN TRANSIT

I have already covered in the history the primary purpose of why this Committee was organized; also, because of the terrible waste of the finished packaged product on the premise that all engineering, manufacturing and quality control efforts are in vain if the product reaches its destination in a damaged condition.

I am not going to dwell too long on this subject as the complete story is shown in the film that I will present later on. I believe the only thing that counts in any kind of a program is results;

therefore, I believe the most interesting information I could give is a few examples received from other manufacturers.

Examples:

Kresky Manufacturing Co., Petaluma, Calif.

Lee Bacon, General Traffic Manager -- "The first part of the NST Program--reduction in transit damages, has been achieved almost 100 per cent by us. Since using the NST procedure, we have not experienced a single instance of shipping damage."

Majestic Manufacturing Co., St. Louis, Mo.

This company reports that their claims for transit damages have been reduced over 50 percent since their supplier redesigned their crate to pass the Safe Transit pre-shipment tests.

Globe American Corp., Kokomo, Ind.

Clark Strayer, Packaging & Methods Engineer -- "Our damage reports prove beyond any question of a doubt that the extra attention given to assembly, packaging, in-plant handling, and carloading methods has paid rich dividends through a reduction in damaged finished products. Our damage results during the period from January 1 through March 30, 1952, were reduced to 1 1/10 per cent. .....We apply the NST label to the exterior of all range crates, and we have found that it has attracted the attention of the carriers, encouraged careful handling, and that it has tended to assure the receivers that our company is doing its utmost to deliver the finished products safely."

The Maytag Company, Newton, Iowa

R. H. Thompson, General Traffic Manager -- "The complete testing equipment recommended by the NST program is employed in our testing laboratory. This pre-shipment testing has proved of great benefit in that it proves our design and method of packing is satisfactory to withstand the hazards of transportation and handling--before shipments are made. We use the NST label, and reports from the field indicate that this program is most helpful."

Norge Division, Borg-Warner Corp., Effingham, Ill.

R. Heine, Chief Inspector -- "Safe Transit has paid for itself many times over in reduced damage costs, dealer and distributor good will, and peace of mind on the part of plant management. When any new development passes our Safe Transit Pre-shipment tests, we are certain it will stand up during transportation."

The Floyd-Wells Co., Boyersford, Pa.

D. Saponara, Plant Engineer -- "Prior to the use of the NST program our company suffered considerable damage in breakage of parts and chipping of enameled surfaces on our ranges, during transit to our dealers.....With the inception of the NST program we very

appreciably reduced both--thereby reducing loss due to replace of parts. Of equal importance was the good will realized by having our ranges reach our dealers in perfect condition. Both are very tangible benefits."

Moffats Limited, Weston, Ontario, Canada

H.W. Bonner, Packaging Engineer-- "At the time of our application for National Safe Transit certification our damages on an overall picture were 24 per cent; this has now been reduced to 2 per cent through NST tests and with the cooperation both our national railroads."

Perfection Stove Co., Cleveland, Ohio

William Day, Chief Engineer -- "We thought the way we were packaging our household ranges was pretty good. But since adopting the Safe Transit pre-shipment testing procedures, we have cut shipping losses from 12 per cent to a negligible percentage."

Geo. D. Roper Corp., Rockford, Ill.

E. H. Shands, Dir. of Engr. and Development -- "We have found by daily testing that small deviations from our dimensional standards, many of which would be of no consequence, would occasionally become particularly troublesome. When this occurs, a daily Safe Transit test detects this trouble immediately, and we are able to correct it which has aided us in reducing our damage in transit to below 1 per cent. The fact that has been particularly surprising to us is that we have reduced the cost of our crating approximately 25 per cent."

THREE

A TOOL FOR REDUCING THE COSTS

By having a definite pre-shipment standard test, it obviously makes it possible to "milk out the last drop" of cost possible and at the same time be sure that the packaged product will reach its final destination in good condition if handled in a normal manner. How is this accomplished? Let me give you a few examples.

Examples:

Day & Night Division, Affiliated Gas Equipment, Inc.,Monrovia,Cal.

V.R. Mottinger, Chief Process Engineer -- "..Another item which makes the program economically sound is the fact that we have been able to review packaging, and have designed new packages at appreciable savings to our company. Savings which have already been effected have more than paid for the testing expenses which have been accumulated."

Tappan Stove Co., Mansfield, Ohio

G. L. Dobson, Chief Production Engineer -- "The standard tests have led us to a better crate design. Fortunately, the new crate which was designed to resist the standard tests turned out to be less expensive than the old one."

Westinghouse Electric Corp., Beaver Works

W. D. Cupps, Suvp. Inspection and Test -- "Dear Mr. Bisbee: Complying with your request of June 25, for the value of what cost reductions have been made due to NST testing of our packaged products, the estimates given are based on six months activity, therefore should be doubled in order to get the savings for a year. Also, there are a number of intangible savings enumerated at the bottom of the estimate that are difficult to evaluate in dollars and cents.

At this time, I would like to express my sincere appreciation to the cooperation you and your personnel have given us, as I believe, as you mentioned in your letter, that it has opened a new field of cost reduction for Beaver operations and the industry." (Mr. Cupps shows detailed estimated savings for 6 months period on packing materials of $3760.45. Also a savings of 34700 lbs. in shipping weights for 6 months period at $1.33 average per hundred at $461.50. The foregoing is just on four styles of safety switch starters. This was the result of just the first few switches analyzed. Before we are through, the direct savings will be multiplied many times.)

Easy Washing Machine Co., Syracuse, N.Y.

James Muirhead -- "I have figures available showing savings of approximately 10% on the cost of a container entirely due to NST pre-shipment testing."

"Pre-shipment testing of the packaged product for a new product and daily production brings out product weak points. For instance, such testing has frequently shown that it is entirely possible to add a small amount to the cost of the product and take twice or three times this amount from the cost of the container. In other words, such a program can be of great assistance to the Product Design Department by helping the design engineer in a proper "design for packaging." This is important, as many have discovered too late, that a new product could have been slightly modified and a much less expensive package could have been used."

Day & Night Division, Affiliated Gas Equipment, Inc., Monrovia, Cal.

V. R. Mottinger -- "Before the NST program was started, our only recourse was to re-design our containers, ship them to various points for return, then examine them to see what had happened. The facilities at the AGA Laboratories for local testing of packaged products has greatly simplified the problem..... We must all realize that packaging is an expensive part of any product. In one case, after the AGA NST tests, we were able to modify the

the crating on one of our water heaters in a manner so that we will realize a saving of about $10,000 a year on this single product.

FOUR

A SALES TOOL

It is obvious from reaping the benefits of the results outlined in One, Two and Three that it is definitely promoting customer good will in being assured that the shipper is doing everything possible to prevent him from having trouble of damaged products that he is purchasing.

An industry-wide NST conference was held at the Palmer House, Chicago, March 16, 1953. The theme of the meeting, "Safe Transit-- A Sales Tool For You."

The attendance was made up of a large number of leading manufacturers, carrier representatives from railroads, railway express, air cargo, American Trucking Associations, leading laboratories of the country and many other associations who are working with us to further this program.

As the main theme of the meeting was "A Sales Tool For You" I am quoting in part a paper by W. H. Murray, Assistant Supervisor of Purchases and Stores of the Georgia Power Company, Atlanta, Georgia.

"We not only buy appliances in carload lots, but we then distribute them to over 100 retail outlets of our own. From these retail stores, we install the merchandise directly in the customers' homes. We have the opportunity to see its condition all along the line, and we also have all of the headaches that arise when merchandise is received damaged.

Not so many years ago, our warehouses over the state were so filled with damaged appliances, and the cost of repairs had become so great, I think some of us seriously doubted the wisdom of an electric utility company continuing in the business of selling electric appliances.

But now - thanks to the National Safe Transit Program - we are beginning to see the flood-tide of damage checked. What this program can do in the years ahead to reduce the terrible losses resulting from damage in transit is dependent entirely on the cooperation it receives from all industry and carriers involved.

No longer can the manufacturer hold his place in the highly competitive market of today who holds to the old ideas that his responsibility ends when he builds a good product. It is also his responsibility to do everything possible to see that it is delivered to the buyer in good condition."---

"The National Safe Transit Program has already helped our company in many ways." -----

"The labels on cartons, placed there by those manufacturers who have qualified under the National Safe Transit Program, which read, 'Make Safe Handling Your Job,' have created a lot of interest among our men. It is our belief that the sight of this label actually results in men all along the line giving a little better care to these particular crates. By using the label, the manufacturer is showing the carriers and his distributors that he is doing everything known to shipping science to see that the product is delivered to its final destination in good condition. We believe the carriers use more care when they know the manufacturer is doing his part." -----

"It goes without saying that the independent distributors are going to think a long time before they purchase merchandise and appliances from those manufacturers who are not sufficiently interested in their product to cooperate in this great program."

"LEADERSHIP IN BUSINESS INHERITS THE RESPONSIBILITY OF MAKING PROGRESS. THE NATIONAL SAFE TRANSIT PROGRAM IS PROGRESS -- LET'S ALL SUPPORT IT !" In any language this spells "a good sales tool."

Because Georgia Power Company really represents the distributor, the dealer down to the customer level, and covers the entire State of Georgia, I believe Mr. Murray's thinking tells a better story on "A Sales Tool for You" than any other means I can think of.

PROGRESS TO DATE

Although the program has been in actual operation about four years, it has been adopted by many of the smallest and largest manufacturers in the industry.

Manufacturers turning out the major volume of goods within our scope are now members -- 120 in number or about 80 per cent of the total volume of goods.

Outstanding laboratories have also been certified so that it is now possible for a manufacturer in any locality to have his products Safe Transit tested. There are now 21 certified laboratories.

CONCLUSION

The National Safe Transit Committee and its program have been sponsored by the Porcelain Enamel Industry as a cooperative plan to reduce shipping damage losses in the major appliances and allied metal industries.

It is now expanded to take in television and radio manufacturers, and later on the scope will be widened to take in many other products. In the interim, many companies not within our scope are using the NST program and are benefiting by the results.

This program provides a new approach to the damage problem, based on the use of pre-shipment test procedures, which, by simulating the hazards of transportation, check the adequacy of the package, the structural resistance of the product design and the carefulness of the processing. These pre-shipment tests have been approved as

representing such hazards by the carriers, the laboratories and the manufacturers.

They have been in use now for over four years and the results achieved with them prove that they do correlate with shipping experience. These tests are now accepted therefore, as scientific, authoritative and accurate.

The benefits that the industry and the carriers have derived from the National Safe Transit Program have been substantial and will increase as it expands further. There has already been a considerable reduction in the total of damage claims on appliances and allied metal products; some manufacturers have found Safe Transit has cut their overall costs; it has helped others to work out better product designs; still others have found that the standardized tests have shown them where they needed better shop control. Relationships between the manufacturers, the carriers and dealers have been reported to have improved in every instance.

The Safe Transit plan is a basic method for reducing shipping damage that can be applied to all products of all industries where shipping damage is a serious problem. But to repeat, three other benefits have been proven:

1. A vital phase of Quality Control,

2. A tool for safely reducing costs,

3. A sales tool.

It is our hope that other industries will see in the principles of Safe Transit the means of solving their own problems. We feel sure that if they apply them, they will be as gratified with the results as have been the major appliance and allied metal industries.

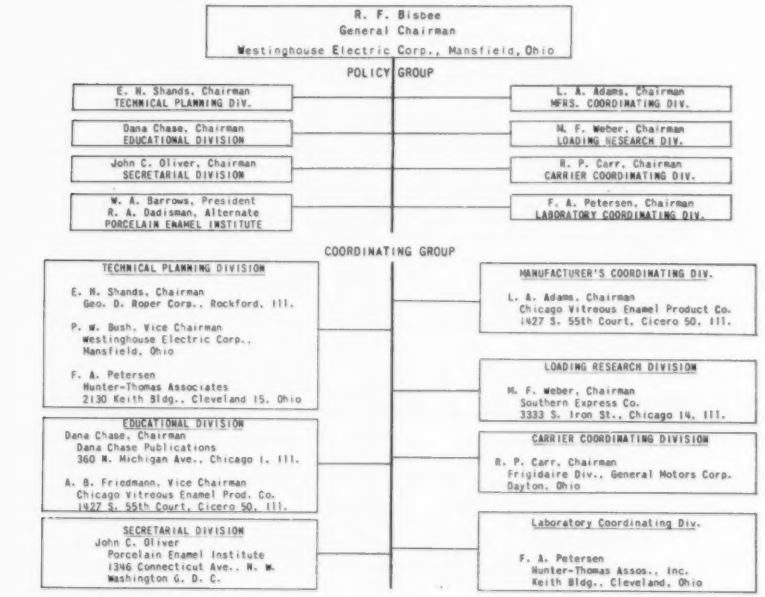
National Safe Transit Film

As this program is now nation-wide in scope, a colored movie has been prepared which we use around the country for various educational meetings. I thought the best way to give you a brief review of these standard tests would be to show you this film, on the basis that a picture tells more than a thousand words and it will be more interesting than just talk.

This film was made by Paul Bush of our Technical Planning Division, and was over two years in the making. Reprints can be obtained for $150.00 from National Safe Transit, Dupont Circle Building, 1346 Connecticut Avenue, N.W., Washington 6, D. C.

This film will conclude my part in this program.

NATIONAL SAFE TRANSIT COMMITTEE



53-215

PROCESS CONTROL IN THE BREWING INDUSTRY BY STATISTICAL METHODS

Robert I. Tenney
Wahl-Henius Institute

Modern methods and practice of producing the clear, amber, foam-topped thirst quencher we all know as beer are directly descended from those of the oldest among the industrial arts. They constitute a brewing technology which combines the best of an ancient art with proven science. A certain conservation accompanies the adoption of each new scientific tool, as in any traditionally based activity. Yet, brewers have been quicker than other industries in utilizing statistical approaches to quality control.

The brewer is perhaps the best qualified to assert the oft-heard phrase, "But we have <u>always</u> had quality control." If we permit "always" to be less than infinite yet stretching beyond recorded data that is certainly true. A great amount of skill has been directed to the control of quality in beverages brewed and fermented from grain since before the dawn of written history.

As mathematically adept as the pyramid builders of the Upper Nile must have been, it is hardly conceivable that they made any practical application of the laws of probability and of frequency distribution curves in purposefully guiding the processing of millet into hek, a forerunner of beer and ale. Yet, before then brewing had become established upon an industrial plane with the responsibilities for control of quality having its own hieroglyphic character. Numerous papyrus descriptions of the many flavors available attest to the variation in quality and the exercise of control. Just as spices and condiments were found useful in covering the malodorous condition of tainted foods so were strongly flavored ingredients mentioned as desirable in maintaining a more uniform and reproducible flavor in ancient beers.

In fact, the use of strong flavors and rigid adherence to empirical procedures remained the only elements of control available for centuries. Chance dictated much of the beer character. Success was measured solely in terms of long runs of favorable influences but was termed "skill". Each fermentation could be caused by any of a host of the different microorganisms in the surrounding air. Selection of the spontaneous origin of fermentation was greatly influenced by chance. Not until Pasteur's work in 1859 and Hansen's studies a few years later were brewers able to purposefully select strains of the more desirable yeasts. Pure cultures have been used in this country since 1892 and have definitely contributed to guided beer quality.

Today, the development of ideal yeast variants calls for the highest statistical skill as the geneticist manipulates a multiplicity of genes in increasing permutations for study and test. The biologist assumes statistical approaches with much greater ease than the chemist or engineer who has been taught definiteness in seeking answers.

The search for better covering flavors continued slowly throughout early history but had culminated in the choice of the delicately aromatic

hop by about 1500. The previous three hundred years had seen a steady increase in the use of hops as compared to rosemary, marjoram, anise and yarrow which had previously been popular. This was a fortuitous choice, for extractives of this plant have rather definite anti-biotic properties and have assisted in maintaining yeast purity.

It was also in this era that the thermometer came into being along with the hydrometer and other devices which were promptly adopted to measure brewery conditions - a first step in any control program. As new tests and measures became available they were turned to use by brewers. Much trading of methods across industries was practiced as an old art was growing into a technology.

Among the brewers' contributions to the general field during this growth we must mention the statistician who wrote under the pen name of "Student" and who gave us the t test for significance. W. S. Gosset, head brewer for the Arthur Guinness Company, makers of fine ale, used this name since company policy at that time was against its employees writing technical papers.

The brewing industry has at this time, no startling contributions to give to the field of Statistical Quality Control. A few of its members are beginning to apply the tools but, in the main, the term "statistical" is just as much of a retardant to brewers as it has been to other process industries. A few papers dealing with statistical methods have been presented in the trade societies' meetings and a first training course was given by Wahl-Henius Institute in January of 1953. This was pitched to management's viewpoint and was attended by management and production men from 21 plants. Quality control is a production responsibility throughout the industry as a matter of tradition. It will take time for it to develop into an exclusive management staff function.

The tools used are the familiar ones and the problems are no different statistically from those of other industries. The brewer is aiming his product at the center area of market preferences which are represented by the frequency distribution curve at the extreme right of Fig. 1. The curve is actually a composite of several curves,

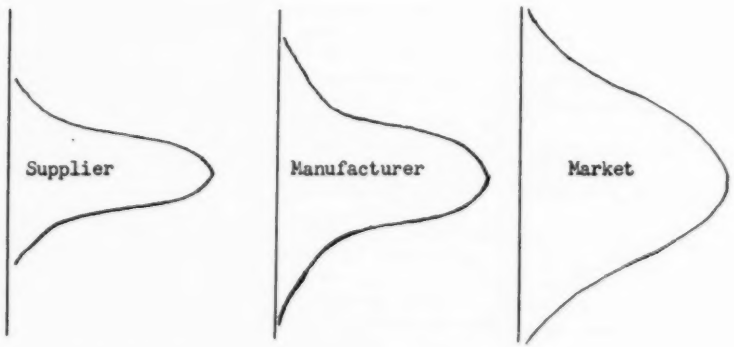


Fig. 1 TARGET RELATIONSHIPS TO AN SQC VIEWER

one for each quality he intends to control. He knows that it is unlikely that his product will completely satisfy, for any single item, much more than a range of preference which is from two to two and a half sigma values wide on this market curve. Some measurable items are even more critical than this. It is of course not difficult to hit a target 6σ wide with a shot gun having a spread of 2σ but the most profitable response is obtained when the mean values are the same. The center curve of Fig. 1, then, represents the brewers possibilities positioned most favorably to the market. The brewer is not unique in this position except perhaps in the availability of wide tolerances for product acceptability providing the market is large enough. Any manufacturer faces a similar situation.

Now, if the brewer, or other manufacturer has analyzed his market in this fashion and recognizes the magnitude of variance within his own product he will inevitably seek raw materials from his suppliers which best enable him to hold his mean on target center. This leads to his requesting rifle shots from his suppliers as indicated by the left curve of Fig. 1.

Flavor is, of course, the brewers number one controlled quality. He does this by balancing several constituents which may be measured separately by chemical or organoleptic methods. Blending is an obvious method of minimizing the batch to batch fluctuations which is followed in most breweries.

The solid lines of Fig. 2 reflect the composition of a hypothetical beer which would have sold the most volume during the twelve months of 1952. The market acceptance curve is dynamic and reflects the total reaction of the public to advertising efforts as well as actual composition and flavor. In compiling the data for these time functional charts a number of different brands of beer have been analyzed each month - enough brands to represent a sample of at least 52% of the total market. The actual value determined for each brand was multiplied by the number of barrels sold of that brand. After adding these products the sum was divided by the total volume of beer sold. Thus the graphs represent the composite of all beers (with reliability of a 52% sample) weighted to sales experience. It will be noted that Apparent Extract (body of the beer) has the least trend toward a change and that sweetness, as reflected by reducing sugar content is rather changeable.

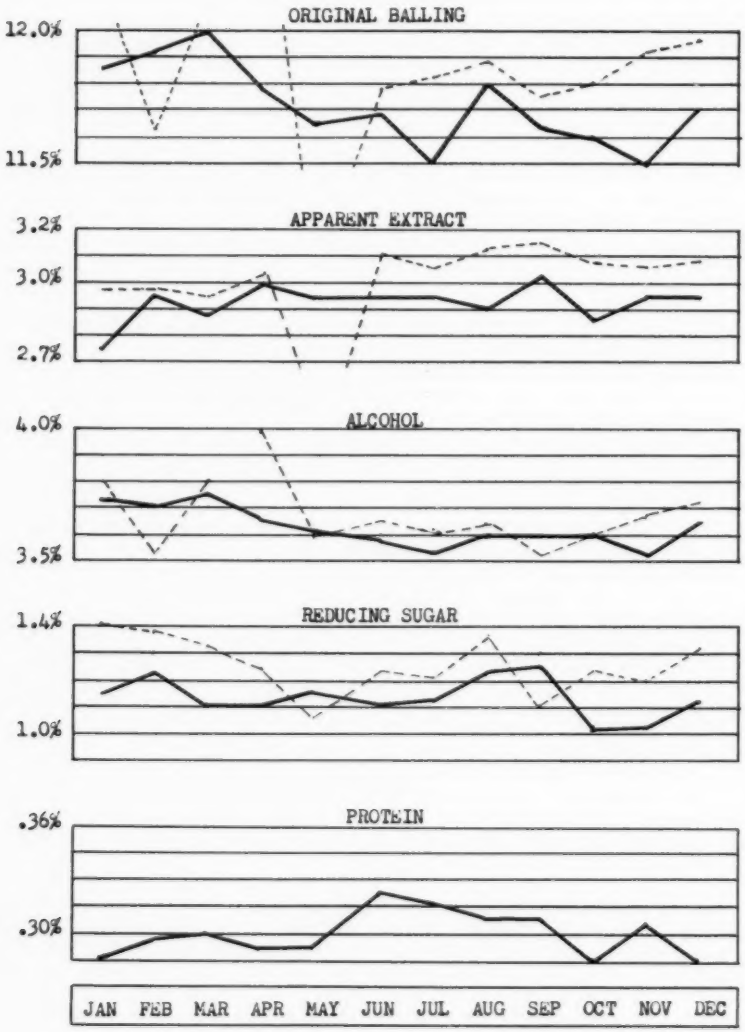
The dotted lines in the graphs of Fig. 2 show values of one typical brand over the same interval. Note how it was blended primarily to uniformity of body with less respect to sweetness or even alcoholic content.

Of course, these graphs have presented but a few of the qualities measured by brewers. Flavor is also guarded by taste testing. In 1410 the town council of Hamburg instituted the "Probestecker". On two days of each week testers examined three tankards of beer from each new brew offered by any of the towns several hundred breweries. If any was found lacking in proper color, taste and body it was confiscated so that the towns reputation for good beer would not be impaired.

The previous paper has dealt with organoleptic tests and some of the statistical approaches to this type of problem. A few others have described the specific problem respecting beer and other beverages and

FIG.2. FLUCTUATIONS IN BEER COMPOSITION

COMPOSITE SAMPLES FOR 1952



Solid Line - Industrial Average

Dotted Line - Single Brand

are mentioned at the end of this article. In most applications a preference rating is sought. A company has been producing a beer or other product in a certain manner and has gained public favor. A proposed change in process is thought desirable but a question arises as to whether it will affect the flavor. Consequently the standard and the proposed products are compared by a company "expert" as to which is preferable and decisions may be based upon that individual's reaction. This is not peculiar to the brewing industry but is typical of how a production superintendent, plant manager or business owner can allow his preference to dictate what he sells the public. In many instances this has brought success to a beer and in many it has brought failure.

The statistically minded brewers realize the fallacy of relying upon one man's judgment and of seeking preference expressions directly. They have instituted regular schedules of panel type testing and seek the ability to differentiate rather than an expression of like or dislike. After all, if one cannot distinguish between two samples how could an honest useable preference exist?

The public is quite label conscious in their tasting and is not apt to be as discerning of beer as are those who make it. Among untrained tasters it is, in fact, difficult to distinguish correctly and at significant levels between a number of brands on today's market. Realizing this the personnel of brewery taste panels are trained in a descriptive vocabulary and in the various elements of flavor perceptions. A number of distinction types of tests are used and the data are evaluated with reference to deviation from that expected from chance distributions. The null hypothesis is a useful tool in this respect.

A common approach consists of comparing three glasses of beer under controlled conditions. Two of the glasses contain one of the samples to be compared and the third contains the second sample. Chance dictates a correct judgment of difference (even if none exists) one time in every three trials. By submitting the samples to at least twenty persons in this fashion one can draw valid conclusions. If fewer judgments must be used then two samples can be adequately compared in five glasses. If three are of sample A and two of B the odds of making a correct choice from chance influences are 1 in 10. The sensitivity of the method is thus increased and the binomial theorem can yield useful figures of significance from any number of tests presented. These methods are adequately described in detail in the references mentioned in the bibliography.

The Control Chart is a familiar device in a number of breweries. It has found widest acceptance in bottling operations where it is useful in helping to prevent over-operation of filling machines and to evaluate bottle washing performance. It has also been of utility in maintaining closer batch to batch uniformity in the brewhouse. The "before and after" picture is entirely that one meets in any industry. Historical data shows deviations due to assignable causes and wide control limits. As the control chart is kept the psychological effect alone helps draw the three sigma limits closer to the central line.

Fig. 3 is typical of a situation that can arise in a brewery without notice and which is more quickly detected under control chart methods that through the recording of data in columnar form. It shows the

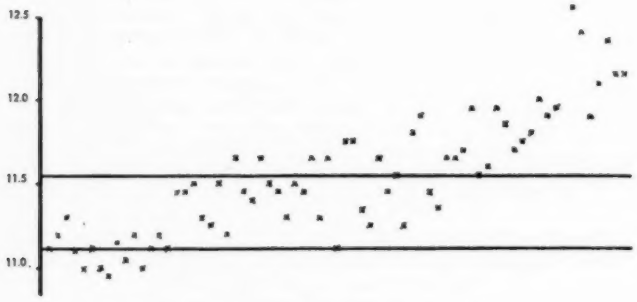


Fig. 3. Wort Extract, Brewery A

extract content of wort from successive brews over a period of several weeks operation. The tabulated data revealed an increase in yield but this was felt desirable since it was better than expected performance.

Had the data been plotted and had the specifications been as well understood as after control charts were adopted the obvious upward trend would have been stopped much earlier. As it was the condition was only forcibly pointed out at the time the statistical program was being adopted. The solid lines indicate the intended range.

Fig. 4 shows data from another plant and reveals differences due to assignable causes which would be corrected faster under a statistical approach.

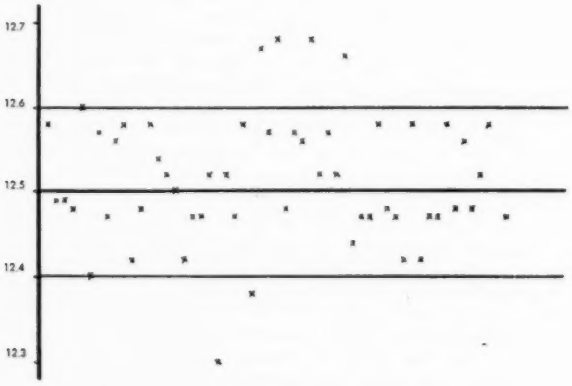


Fig. 4. Wort Extract, Brewery B

Vendor ratings in a true statistical sense have not been widely used by brewers. At least three but fewer than six of the countries approximatelely 250 breweries have partial programs in "this" useful SQC field. These are usually applied by constructing frequency distribution

diagrams of a large number of lots submitted by each supplier of variable items such as malt and adjuncts. Penalty points are then awarded upon a Standard Deviation scale. Care must be taken not to penalize twice for correlated measurements and full discussion between vendor and buyer are necessary before agreeing upon penalty values.

Due to the natural deviation of malt and other major brewing materials acceptance sampling methods are not advisable for brewers without previous discussions with suppliers. When specifications exist, tests upon a single sample taken in proscribed manners are sufficient for rejection. At times a referee laboratory may be called upon in a controversy. Acceptance sampling plans have, however, been useful in evaluating shipments of manufactured items such as bottles, cartons, crowns, bungs and the like. A statistical approach to bottle purchasing has been pushed for several years by the Brewers Industrial Foundation. As yet the brewers desired specifications are narrower than the inherent variations of manufacture due to mold wear and other factors which the bottle manufacturers have already subjected to statistical techniques for improvement. This is a typical example of the situation depicted in Fig. 1.

Acceptance sampling with an aspect of control using number of defects may become useful in supplementing if not supplanting the 100% inspection now employed by all brewers. Each bottle is examined at least once and frequently as often as three times in its trip from filler to case packer in normal operations. The desire for perfect eliminations of all bottles containing foreign bodies is so high that it appears unlikely that statistical techniques will eliminate the practice of 100% inspection.

Statistical Analysis of data with particular reference to the laws of probability have been a help to many brewers in refining their quality control programs. It represents a manner of thinking and properly recognizes the inevitableness of variation. It leads to the ability to predetermine the quality level of product to best fit the market desires.

BIBLIOGRAPHY

Bengtsson, K. & Helm, E. - "Principles of Taste Testing" Wallerstein Laboratory Communications 9 171 (1946)

Mayer, Morton, & Laufer - "Application of Control Charts to Evaluation of Brewhouse Yields" American Chemical Society Meeting, Milwaukee April, 1952

Peryam, D. R. - "Measurement & Control of Flavor Quality" 4th Annual Convention A.S.Q.C. Milwaukee, 1950

Swain, R. C. & Bell, V. P. - "The Flavor Profile ... A Method for Measuring Beer "Taste Appeal"- Modern Brewery Age Nos. 6 & 7, 1951

Tenney, R. I. - "Statistical Quality Control Methods - An Introduction to Brewery Application" - Proceedings American Society Brewing Chemists, 1952

Tenney, R. I. - "A Statistical Approach to Quality Control" Proceedings Master Brewers Association of America, 1952

Tenney, R. I. - "More on Flavor - Control" Communications Master Brewers Association 12 Nos. 1 & 2 (1951)

Tenney, R. I. - "In Beer Character - What Does the Public Want?" Modern Brewery Age, December 1951

VonMeyer, W. H. - "The Use of $\bar{X}$ and R Charts in Evaluation of the Efficiency of Two or More Bottle Washing Machines" Proceedings American Society Brewing Chemists, 1952
Virden, C. J. - "The Design and Interpretation of Tasting Experiments" Journal Industrial Brewing 55, 228-233 (1949)

STATISTICAL PROBLEMS IN LIFE TESTING[1]

Benjamin Epstein
Wayne University

I. Introduction

It is my object in this paper to describe some recent research on statistical problems in life and fatigue testing. It seems appropriate that this topic be discussed before the American Society for Quality Control, which has shown from its beginning a keen and active interest in new statistical developments, particularly as they relate to possible practical applications. The topic that we shall discuss has within it aspects which should be of considerable interest to both the theoretical statistician and the man faced with the practical problem of making decisions on the basis of life test data.

What we shall say in this paper can be considered a contribution to and application of the theory of order statistics. It is a characteristic feature of life test and fatigue data that the information becomes available in an ordered manner. In most problems that we are acquainted with data do not become available in order. Indeed, if ten items are taken from a machine and measured for some characteristic such as length, or weight, or diameter it would be considered evidence of lack of control to have the observed magnitudes come to us in either strictly ascending or descending order. However, this is precisely what does (and must) happen in life testing, fatigue testing, and in a variety of destructive test situations. If, for example, n radio tubes are put on a life test, then the weakest one will fail first in time, the second weakest one will fail next, etc. Exactly the same kind of ordered situation will occur whether we are concerned with the life of electric bulbs, life of radio tubes, life of ball bearings, or the length of life of human beings or animals after undergoing treatment for a disease. Order need not necessarily be order in time. If we are interested in such things as the current needed to blow a fuse, the voltage needed to break down a condenser, the pressure needed to break a bottle, or the force needed to rupture a physical material, then we can, if we wish, arrange to carry out testing in such a way that every item in the sample is subjected to precisely the same stimulus (current, voltage, pressure, stress, etc.). If this is done, then clearly the weakest item will be observed to fail first, the second weakest next, etc.

Put in general terms then, we envisage a situation where n items are drawn at random from some population and where the data become available in order of magnitude. Clearly we can, if we choose, discontinue experimentation long before all n items have actually failed. In particular we may decide to truncate the experiment as soon as we have the first r ($\leqslant$ n) failures, or we may decide to stop experimentation at some preassigned time of truncation T_o, or we may decide to follow some sequential procedure permitting us to stop as soon as certain conditions are met. In all of these situations our primary concern is to develop statistical procedures, which by taking advantage of the fact that data

(1) Research sponsored by the Office of Naval Research.

become available in order, will enable the experimenter to reach a decision in a shorter time or with fewer observations or both, than would be possible if one were not given ordered information.

II. The Exponential Distribution and its Role in Life Testing

From the beginning, the work that we have been doing concerns itself in large measure with seeing how one can best utilize the information contained in the first r out of n failures if the characteristic X being investigated has an exponential distribution described by a probability density function $f(x;\theta)$ of the form

$$(1) \qquad f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0, \ \theta > 0.$$

If x is considered as life in hours it appears that by choosing θ suitably one can get a reasonable first approximation to the life of electron tubes by using the density (1). Before going into some of the results obtained when the density function is exponential, it seems worthwhile to make a few remarks about the validity of the exponential distribution as a model. We have made an effort to find out as much as we could about actual experience in a variety of applications. While the information that we have been able to get is fragmentary and inconclusive, it is interesting and pertinent to the present discussion.

Before discussing specific references we should like to remark that there is a basic theoretical reason why the exponential distribution of life should play an important role in some applications. As pointed out by Feller in chapter 17 of [1] a fundamental property characterizing the exponential distribution is the following (stated in life testing language): The probability that an item will live until time (T + t) given that it has lived until time T is independent of T. The only continuous distribution with this property is the exponential. The discrete analogue of this result is that the geometric distribution is uniquely characterized by the fact that, if in a sequence of independent Bernoulli trials k successive failures have occurred, then the additional waiting time (measured in number of additional trials) required to observe a success for the first time is independent of k.

An immediate consequence of having a life distribution characterized by a density function which is exponential is that the rate of failure must be constant.

Let us now return to the main stream of our discussion and make some remarks on literature bearing on the validity of the exponential assumption. The most recent and extensive reference is a paper by Davis [2]. In this paper Davis gave several examples which lend substantial support to the view that actual life data drawn from a variety of fields follow in the main either an exponential or a normal distribution. As a principal example of exponential life distributions, Davis gave data based on observations made on a number of types of electron tubes. As an example of a normal life distribution, he gave the case of incandescent lamps. An interesting mixed example is furnished by bus motor failures, where apparently first failures (measured in miles) are fairly normal, while third, fourth, fifth, and higher failures (measured in miles between appropriate successive failures) are fairly exponential.

Another reference is a report [3] issued under the George Washington University Logistics Project. One of the principal findings was that data obtained from tube failures under certain operational conditions justified the assumption that the underlying life distribution was exponential. Various excellent fits to actual data were obtained. Another report [4] of interest in this connection deals with tube life experience on one of the M.I.T. High Speed Computers. Tube life experience is in reasonable accord with an underlying life distribution which is exponential. It is interesting to quote from page 4 of [4] : "A question which is sometimes raised in discussions is whether tubes tend to have a definite period of usefulness and therefore should be replaced as a group after a certain length of time or whether the failures occur at a fixed rate so that nothing would be gained by replacement. Our experience with the multiplier seems to point to the latter type of behavior and certainly indicates that wholesale replacement is not warranted within the period of time shown on these curves." (Figs. 1, 2, and 3 of [4]). This is a practical way of saying that the assumption of a constant rate of failure is completely equivalent to assuming an exponential distribution of life.

In [3] there is a rather complete bibliography on the subject of industrial replacement. One of the references [5] is to life test data on valves (British terminology for electron tubes). The exponential distribution seems to be a good fit for the data obtained from the life time to failure of 700 valves. Two other and somewhat older references are [6,7] .

Although the evidence just cited would seem to strongly suggest that the exponential distribution of life is a reasonable assumption to make, it is necessary to emphasize at this point that there is other evidence which suggests that there is more to the problem than meets the eye. E.g., in [7] , the exponential probability density function of life was obtained after eliminating a large number of early failures due to all kinds of mechanical failures, shorts, open circuits, etc. The existence of two kinds of failures is also emphasized in recent papers by McElwee and others [8, 9] and in a series of unpublished conference papers by Steen, Wilde, and Davis [10, 11, 12, 13] presented at the Summer Statistics Conference sponsored by the University of North Carolina in July, 1952. In all of these papers there seems to be common agreement on the existence of essentially two kinds of failures, one of which occurs either at the outset of testing or within the first few hundred hours. The other kind is a very gradual, long drawn out kind of failure due to the "dying away" of certain performance characteristics. There also seems to be fairly general agreement that if the first kind of failure can be substantially eliminated, then it would be reasonable to assume a constant rate of failure situation (i.e. exponential life). It might be added parenthetically that if early failures are as frequent as reported, then it might be highly desirable in most applications to eliminate the majority of such tubes by some kind of burning in period lasting for a number of hours dictated by previous experience.

III. Some Statistical Results on Exponential Life Test Data.

It should be fairly clear from the discussion in section II (notwithstanding the remarks in the last paragraph of that section) that it is of considerable interest to see what results can be obtained

when the underlying density of life is given by $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$ (2). We will not prove any of these results here. The reader who is interested in the mathematical derivations and other details is referred to several reports and papers [14, 15, 16, 17, 18, 19]. The first result is as follows: Let n items be drawn at random from a distribution whose density is given by (1). Let the observations become available in order, i.e., $x_{1,n} \leq x_{2,n} \leq \ldots \leq x_{r,n} \leq \ldots \leq x_{n,n}$ (where by $x_{i,n}$ is meant the i'th smallest observation in a sample of n ordered observations). Suppose experimentation is discontinued after the first r observations are made (r is preassigned), then it can be shown that the maximum likelihood estimate of the mean life θ is given by $\hat{\theta}_{r,n}$ where

$$\hat{\theta}_{r,n} = \frac{x_{1,n} + x_{2,n} + \ldots + x_{r,n} + (n-r)\, x_{r,n}}{r} \quad (2)$$

The estimate $\hat{\theta}_{r,n}$ is "best" in the sense that in addition to being maximum likelihood, it is also unbiased, minimum variance, efficient, and sufficient. The probability density function of $\hat{\theta}_{r,n}$ is given by

$$f_r(y) = \frac{1}{(r-1)!} (r/\theta)^r\, y^{r-1}\, e^{-ry/\theta}, \quad y > 0 \quad (3)$$

$$= 0, \text{ elsewhere}$$

and $2r\hat{\theta}_{r,n}/\theta$ is distributed as chi-square with 2r degrees of freedom (which we denote as $\chi^2(2r)$).

In the preceding paragraph we have been concerned with the non-replacement situation where one does not replace failed items at once by new items drawn from the underlying density function (1). In the replacement case it can be shown that the maximum likelihood estimate of the mean life θ is given by

$$\hat{\theta}_{r,n} = n\, x_{r,n}/r\,, \quad (4)$$

where by $x_{r,n}$ is meant the total time (measured from the beginning of the experiment) to observe the r'th failure. n is the sample size which is maintained the same throughout the experiment. The estimate (4) enjoys exactly the same optimum properties in the replacement case

---

(2) Most of the results achieved for the density function $\frac{1}{\theta} e^{-x/\theta}$ also hold for the cumulative distribution

$$F(x) = 1 - e^{-g(x)/\theta}, \quad x \geq 0 \quad (\theta > 0)$$

$$= 0, \quad \text{otherwise},$$

where $g(x)$ is a strictly increasing function of x for $x \geq 0$ with $g(0)=0$ and $g(\infty) = \infty$. For example the maximum likelihood estimate of θ is then given by

$$\hat{\theta}_{r,n} = \frac{g(x_1) + \ldots + g(x_r) + (n-r)\, g(x_r)}{r}.$$

As a case in point, we may mention the Weibull distribution where $g(x) = x^b$ (b a known constant).

as does the estimate (2) in the non-replacement case. Furthermore the probability density function of the estimate (4) is given by (3).

An interesting and important feature of the distribution of $\hat{\theta}_{r,n}$, in either the replacement or the non-replacement case, is the fact that it is independent of n. From the point of view of estimation this means that we have a whole class of equivalent procedures for estimating the mean life θ. For each member of the class $E(\hat{\theta}_{r,n}) = \theta$ since the estimate is unbiased. Further $\mathrm{Var}(\hat{\theta}_{r,n}) = \theta^2/r$. From the point of view of acceptance testing we can exhibit a whole class of equivalent test procedures for deciding whether to accept some high (and presumably desirable) mean life θ_o and to reject some low (and presumably undesirable) mean life θ_1 with preassigned producer's and consumer's risk. It can be shown that the region of acceptance for $\theta=\theta_o$ must be of the form $\hat{\theta}_{r,n} > C$ and further the operating characteristic curve based on this region of acceptance must be the same for all n. Details for finding the r and C appropriate to a given θ_o, θ_1, α, and β are given in [14, 18]. We shall now discuss briefly an interesting feature of the class of tests and estimation procedures based on using the first r out of n ordered observations.

For definiteness let us compare two non-replacement procedures: Procedure 1 involves basing decisions and estimates on the first r out of n ordered observations and Procedure 2 involves doing similar things when n=r (i.e., one takes all r out of r observations). Since these procedures are exactly equivalent in the sense that $\hat{\theta}_{r,r}$ and $\hat{\theta}_{r,n}$ have the same distribution, our only justification for preferring Procedure 1 to Procedure 2 is that it will take a shorter time on the average.(3) For example a test procedure which involves taking the smaller of two random observations will lead to tests whose operating characteristic curve is exactly the same as that found from observing 1 out of 1. However, the expected length of time in the first procedure is only one-half that in the second procedure. Consequently, the procedure involving the use of the smaller of two observations and stopping there is to be preferred to the one involving taking just one observation at

(3) From the point of view of the number of items destroyed, both procedures involve the destruction of r items. In fact it might be argued that a procedure which is based on using the first r out of n ordered observations really uses n items. After all it might be said that the (n-r) items which have not failed are not as good as new. To this kind of objection there are at least two answers:

(i) According to our remarks in Section II , it is theoretically correct to say that the (n-r) remaining items are as good as new. This follows from the exponential property.

(ii) Even if the (n-r) items are not as good as new, it is often of critical importance to be able to come to a decision quickly. The decision might involve the disposition of thousands of items and the probability of coming to a decision quickly without increasing the risks of making a wrong decision might well be worth the cost of (n-r) additional items.

random, if the saving in time outweighs the loss due to testing two items rather than one.

In the non-replacement case it can be shown [14, 18] that $E(X_{r,n})$, the expected length of time needed to observe the first r out of n ordered observations is given by

$$(5)\quad E(X_{r,n}) = \theta\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-r+1}\right) = \theta \sum_{j=1}^{r} 1/(n-j+1) .$$

If r=n, (5) becomes

$$(6)\quad E(X_{r,n}) = \theta\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{r}\right) = \theta \sum_{j=1}^{r} 1/j .$$

In the replacement case one gets

$$(7)\quad E(X_{r,n}) = r\theta/n .$$

In the non-replacement case let us define $\alpha_{r,n}$ as

$$(8)\quad \alpha_{r,n} = E(X_{r,n}) \,/\, E(X_{r,r}) .$$

The ratio $\alpha_{r,n}$ is clearly a measure of the expected saving of time due to using Procedure 1 (first r out of n) as compared with the Procedure 2. In Table 1 we give the values of this ratio for selected small values of r and n. This table indicates that if "time is money", procedures which use ordered observations may be very advantageous. In [14] tables of $E(X_{r,n})$ and $\alpha_{r,n}$ are given for r=1(1) n and n=1(1)20(5)30(10)100. As a numerical example the reader can verify that for r=2, n=4, one gets $\alpha_{2,4} = 7/18$.

Table I

Ratio of the Expected Waiting Time to Observe the r'th Failure in Samples of Size n and r respectively.

$$E(X_{r,n}) \,/\, E(X_{r,r}) = \alpha_{r,n}$$

r \ n	1	2	3	4	5	10	15	20
1	1	.50	.33	.25	.20	.10	.067	.050
2	-	1	.56	.39	.30	.14	.092	.068
3	-	-	1	.59	.43	.18	.12	.087
4	-	-	-	1	.62	.23	.14	.104
5	-	-	-	-	1	.28	.18	.125
10	-	-	-	-	-	1	.35	.23

IV Truncated Procedures.

In either the replacement or non-replacement situation one may work with truncated tests in which it is decided in advance that the experiment will be terminated at $\min(X_{r_o,n}, T_o)$, where $X_{r_o,n}$ is the time at which the r_o'th failure occurs and T_o is the truncation time beyond which the experiment will not be run. Both r_o and T_o are preassigned in advance. If the experiment is terminated at $X_{r_o,n}$ (i.e., r_o failures occur before time T_o) then the action taken will be to reject. If the experiment is terminated at time T_o (i.e., the r_o'th failure occurs after time T_o) then the action in terms of hypothesis testing is acceptance. In [17] we give details concerning such test procedures for both the replacement and non-replacement cases. Interesting features of these tests are $E_\theta(r)$, $E_\theta(T)$, and $L(\theta)$, i.e., the expected number of observations to reach a decision, expected waiting time to reach a decision, and probability of accepting, if θ is the true value.

In the non-replacement case

$$(9)\quad E_\theta(r) = np_\theta\left[\sum_{k=0}^{r_o-2} b(k;n-1,p_\theta)\right] + r_o\left[1 - \sum_{k=0}^{r_o-1} b(k;n,p_\theta)\right]$$

where $p_\theta = 1 - e^{-T_o/\theta}$ and $b(k;n,p_\theta) = \binom{n}{k} p_\theta^{\,k}\,(1-p_\theta)^{n-k}$.

As a matter of fact the probability distribution of r is given by

$$(10)\quad \Pr(r=k|\theta) = b(k;n,p_\theta),\ k=0,1,2,\ldots,r_o-1$$

and

$$(10')\quad \Pr(r=r_o|\theta) = 1 - \sum_{k=0}^{r_o-1} \Pr(r=k|\theta)\ .$$

Further one has

$$(11)\quad E_\theta(T) = \sum_{k=1}^{r_o} \Pr(r=k|\theta)\, E_\theta(X_{k,n})\ ,$$

where $E_\theta(X_{k,n})$ is given by (5), and

$$(12)\quad L(\theta) = \sum_{k=0}^{r_o-1} \Pr(r=k|\theta)\ .$$

In the replacement case the probability distribution of r is given by

$$(13)\quad \Pr(r=k|\theta) = p(k;\lambda_\theta)\ ,\ k=0,1,2,\ldots,r_o-1$$

and

$$(13')\quad \Pr(r=r_o|\theta) = 1 - \sum_{k=0}^{r_o-1} p(k;\lambda_\theta).$$

In (13) and (13'), $\lambda_\theta = nT_o/\theta$ and $p(k;\lambda_\theta) = \lambda_\theta^{\,k} e^{-\lambda_\theta}/k!$.

Further one has

$$(14)\quad E_\theta(r) = \lambda_\theta\left[\sum_{k=0}^{r_o-2} p(k;\lambda_\theta)\right] + r_o\left[1 - \sum_{k=0}^{r_o-1} p(k;\lambda_\theta)\right]$$

$$(15)\quad E_\theta(T) = \theta E_\theta(r)/n$$

and

$$(16)\quad L(\theta) = \sum_{k=0}^{r_o-1} p(k;\lambda_\theta).$$

In section III, we discussed an acceptance procedure based on the first r out of n ordered observations and stated without proof that it is of the form $\hat{\theta}_{r,n} > C$. In the replacement case this test is clearly truncated, with truncation time rC/n, since according to (4) $\hat{\theta}_{r,n} = nx_{r,n}/r$. In the non-replacement case we are no longer dealing with a situation in which truncation occurs at a fixed time T_o. However, if one rewrites $r\hat{\theta}_{r,n}$ as

$$(17)\quad r\hat{\theta}_{r,n} = nx_{1,n} + (n-1)(x_{2,n}-x_{1,n}) + \ldots + (n-r+1)(x_{r,n}-x_{r-1,n}),$$

it is clear that since every member on the right is non-negative, it may be possible to terminate experimentation before r failures occur. Indeed it can be shown that if the test procedure is $\hat{\theta}_{r,n} > C$, then the probability of terminating with a decision involving ρ failures is given by

$$(18)\quad \Pr(\rho=k|\theta) = p(k;\mu_\theta),\ k = 0, 1, 2, \ldots, r-1$$

and

$$(18')\quad \Pr(\rho=r|\theta) = 1 - \sum_{k=0}^{r-1} \Pr(\rho=k|\theta)$$

where $\mu_\theta = rC/\theta$ and $p(k;\mu_\theta) = \mu_\theta e^{-\mu_\theta}/k!$.

Further,

$$(19)\quad E_\theta(\rho) = \mu_\theta\left[\sum_{k=0}^{r-2} p(k;\mu_\theta)\right] + r\left[1 - \sum_{k=0}^{r-1} p(k;\mu_\theta)\right].$$

$$(20)\quad E_\theta(T) = \sum_{k=1}^{r} \Pr(\rho=k|\theta)\, E_\theta(X_{k,n})$$

where $\Pr(\rho=k|\theta)$ is given by (18) and (18') and $E_\theta(X_{k,n})$ is given by (5).

Finally

$$(21)\quad L(\theta) = \sum_{k=0}^{r-1} p(k;\mu_\theta).$$

The ideas just discussed are very closely connected with the so called G. E. Total Life Plan [13] and (17) is in fact the total life observed up to and including the r'th failure time. One of the unpleasant features of the G. E. plan is the difficulty of computing its operating characteristic curve. The truncated plans discussed in this section have the same general objective as the G. E. plan, and have the further advantage of being capable of theoretical treatment.

V. A Numerical Example.

As an illustration of the theory in Section IV we consider three test procedures which have virtually the same operating characteristic curve. Specifically, it is assumed that we wish to test $H_o:\theta_o = 1500$ hours against $H_1:\theta_1 = 500$ hours, with $\alpha = \beta = .05$; i.e. we want $L(\theta_o) = 1 - \alpha = .95$ and $L(\theta_1) = \beta = .05$ (actually we have to be satisfied with $L(\theta_1) \leq .05$). The three procedures are:

(A) 20 items are taken at random from the lot and placed on life test. Items which fail are not replaced. At each moment t, compute the total life

$$\sum_{k=1}^{i} (n-k+1)\, x_{k,n} + (n-i)\,(t-x_{i,n}),$$ where i is the number of failures which have occurred before time t and n = 20 (if i=0, define total life as nt). If this sum exceeds 8150 for any i such that $0 \leq i \leq 9$, stop the experiment at time t and accept H_o ($\theta_o = 1500$ hours). Otherwise the action taken is to reject. This test is equivalent to accepting H_o if $\hat{\theta}_{10,20} > 815$ and rejecting H_o if $\hat{\theta}_{10,20} < 815$.

(B) 20 items are taken at random from the lot and placed on test. Failed items are not replaced. If $\min\left[X_{10,20}, 540\right] = 540$ (i.e., the tenth failure occurs after 540 hours) truncate the experiment at 540 hours with acceptance of H_o. If $\min\left[X_{10,20}, 540\right] = X_{10,20}$ (i.e. the tenth failure occurs before 540 hours) truncate the experiment at $X_{10,20}$ with the rejection of H_o.

(C) 20 items are taken at random from the lot and placed on test. An item which fails is replaced at once by a new item from the original lot. The time $X_{i,n}$ when the i'th failure occurs is measured from the beginning of experimentation. If $\min\left[X_{10,20}, 407.5\right] = 407.5$, truncate the experiment at 407.5 hours with the acceptance of H_o. If $\min\left[X_{10,20}, 407.5\right] = X_{10,20}$, truncate the experiment at $X_{10,20}$ with rejection of H_o.

In Table 2 we give $L(\theta)$, $E_\theta(r)$, and $E_\theta(T)$ for the test A, B, and C for selected values of θ.

TABLE 2

Properties of Three Test Procedures

Mean Life θ	$L(\theta)$			$E_\theta(r)$			$E_\theta(T)$		
	A	B	C	A	B	C	A	B	C
250	.000	.000	.000	10	10	10	167.2	167.2	125.0
500	.038	.043	.038	9.93	9.94	9.93	331.4	331.6	248.3
750	.355	.366	.355	9.10	9.25	9.10	444.7	453.5	341.3
1000	.698	.702	.698	7.68	8.06	7.68	481.8	509.1	384.0
1250	.876	.877	.876	6.39	6.93	6.39	484.8	529.2	399.3
1500	.950	.950	.950	5.39	6.02	5.39	474.7	536.0	404.5
1750	.979	.979	.979	4.64	5.30	4.64	466.0	538.3	406.3
2000	.991	.991	.991	4.07	4.73	4.07	458.3	539.4	407.0
2250	.996	.996	.996	3.62	4.27	3.62	452.3	539.7	407.3
2500	.998	.998	.998	3.26	3.88	3.26	447.3	539.9	407.4

VI. Sequential Life Tests

One can make substantial improvements on the procedures treated thus far by following a sequential procedure. This point was illustrated in [20] by comparing the expected time, $E_\theta(T)$, and expected number of items destroyed, $E_\theta(r)$, for seven different test procedures. Without going into details we assert that one can apply the sequential probability ratio test of A. Wald [21] to life testing. The interesting point now is that decisions can now be made continuously in time[4]. At each moment t, one can decide either to accept, to reject, or to continue experimentation. It can be shown, in the exponential case, when one is testing H_o: $\theta=\theta_o$ against H_1: $\theta=\theta_1$, that the decision as time unfolds depends on

$$(22) \qquad B < \left(\frac{\theta_o}{\theta_1}\right)^r e^{-\left(\frac{1}{\theta_1}-\frac{1}{\theta_o}\right) L(t)} < A,$$

(4) The concept of continuous sequential decision procedures has been considered independently in an interesting report by Dvoretsky, Kiefer, and Wolfowitz [22]. There are interesting connections between their work and the work that is being done in life-testing.

where $B = \frac{\beta}{1-\alpha}$, $A = \frac{1-\beta}{\alpha}$, and r is the number of failures observed by time t. The decision to continue experimentation is made so long as the inequality (22) holds. As soon as (22) is violated, one accepts H_o if the function of t in (22) is < B, and one rejects H_o (accepts H_1) if the function of t in (22) is > A. In the non-replacement case[5]

$$(23)\quad L(t) = nx_{1,n} + (n-1)(x_{2,n}-x_{1,n}) + \ldots + (n-r+1)(x_{r,n}-x_{r-1,n}) + (n-r)(t-x_{r,n}),\ \text{if } 1 \leqslant r \leqslant n$$

and

$$= nt, \qquad \text{if } r = 0.$$

In the replacement case

$$(24)\qquad L(t) = nt.$$

For example, if $\theta_o = 1500$, $\theta_1 = 500$, $\alpha = \beta = .05$, (22) becomes

$$(25)\qquad \frac{1}{19} < 3^r e^{-L(t)/750} < 19.$$

Suppose, for instance, that the experiment evolves in such a way that we keep on waiting without observing a failure. How long should we wait before making the decision to accept H_o ($\theta_o = 1500$, the higher mean life)? According to (25) this means as soon as $e^{-L(t)/750} < 1/19$. Assuming that the initial sample size is 20, it is clear that $t = \frac{750 \log 19}{20} = 110$ hours. Thus if no failure occurs by the time 110 hours have elapsed, we would stop experimentation with the acceptance of H_o as the true state of affairs. On the other end of the scale, let us suppose that observations are occurring in rapid succession so that effectively L(t) is near zero. In this case the inequality $3^r > 19$ is of interest. The smallest integer r for which this inequality holds is r = 3. Thus if 3 failures occur very early in the test, one would stop experimentation with the rejection of the hypothesis of a high mean life. It should be clear that the decision to accept H_o, the higher mean life, can be made continuously. If one waits too long between $x_{r,n}$ (the r'th failure time) and $x_{r+1,n}$ (the (r+1)'st failure time), one should stop experimentation somewhere in between. On the other hand, the decision to reject H_o can only be made right after a failure has occurred. The procedure was accordingly termed semi-continuous in [20].

A good deal of theoretical work which has been done in connection with this problem, will appear in detail elsewhere. We shall be content

(5) It should be noted that the concept of total life occurs in a natural way here. It should also be remarked that in the non-replacement case a special problem arises if all n items are used up without reaching a decision. This point can be taken care of in various ways.

here merely to compare replacement procedure C in section V with a continuous sequential replacement procedure which has the same operating characteristic curve as C. The comparison is given in Table 3.

TABLE 3

Comparison of Replacement Procedure C with a Sequential Replacement Procedure (S):

$$\theta_o = 1500,\ \theta_1 = 500,\ \alpha = \beta = .05.$$

Test Procedure	Expected number of hours to come to a decision		Expected number of items destroyed	
	If H_o is true	If H_1 is true	If H_o is true	If H_1 is true
C	404.5	248.3	5.39	9.93
S	220.5	153.5	2.94	6.14

VII. Conclusion

The results mentioned briefly in this talk should indicate at least the direction of our work. Much more has been and is being done on theoretical problems in this field by several people [23, 24] including the author.

References

1. W. Feller, "Probability Theory and Its Applications", John Wiley and Sons, 1950.

2. D. J. Davis, "An Analysis of Some Failure Data", Journal of the American Statistical Association, 47, 113-150, 1952.

3. C. R. M. Tuttle and A. R. Frank, "Inventory Control Methods Applied to Electronic Tubes". This a paper included in Logistics Papers, Issue 8, 16 November, 1951 - 15 February, 1952, Appendix I and issued by the George Washington University Logistics Project.

4. E. S. Rich, "Experience with Receiving-Type Vacuum Tubes on the Whirlwind Computer Project", Report R-194, February, 1951, Project Whirlwind, Servo-mechanisms Laboratory, M.I.T.

5. N. W. Lewis, "Note on the Exponential Distribution in Statistics", Post Office Electrical Engineers' Journal, 41, 10-12, 1948.

6. D. K. Gannett, "Determination of the Average Life of Vacuum Tubes", Bell Tel. Lab. Rec., 378-381, 1940.

7. J. E. M. Lagasse and W. H. Dean, "Tubes at Work", Electronics, 56-62, 1941.

8. M. A. Acheson and E. M. McElwee, "Concerning the Reliability of Electron Tubes", the Sylvania Technologist, 4, April, 1951.

9. A. J. Heitner and A. L. Wilson, "Analysis of Kew Gardens Premium Tube Life", Sylvania Electric Products, Inc., 83-30 Kew Gardens Rd., Kew Gardens, 15, L.I., N.Y., May, 1951.

10. J. R. Steen, "Life Testing of Electronic Tubes".

11. R. D. Wilde, "Nature of Rate of Failure Curves".

12. R. D. Wilde, "An Example of a Statistically Designed Experiment in a Radio Tube Manufacturing Process".

13. J. A. Davies, "Life Test Acceptance Sampling Methods". (References 10, 11, 12, 13 are to papers delivered at a Summer Statistics Conference sponsored by the University of North Carolina in July, 1952).

14. B. Epstein and M. Sobel, "Some Tests Based on the First r Ordered Observations Drawn from an Exponential Distribution", Wayne University Technical Report No. 1 and Stanford University Report No. 6, March, 1952.

15. B. Epstein, "Estimates of Mean Life Based on the r'th Smallest Value in a Sample of Size n Drawn from an Exponential Distribution", Wayne University Technical Report No. 2, July 1, 1952.

16. B. Epstein and C. K. Tsao, "Some Two Sample Tests Based on Ordered Observations from the Exponential Distribution", Wayne University Technical Report No. 3, December 15, 1952. Also submitted for publication.

17. B. Epstein, "Some Results on Truncated Life Tests in the Exponential Case", Wayne University Technical Report No. 4, February 15, 1953. Also submitted for publication.

18. B. Epstein and M. Sobel, "Life Testing", accepted for publication by the Journal of the American Statistical Association.

19. B. Epstein and M. Sobel, "Some Theorems Relevant to Life Testing from an Exponential Distribution". Submitted for publication.

20. B. Epstein and M. Sobel, "Life Testing". A lecture delivered at the Stanford Sampling Inspection Conference, August 21, 1951.

21. A. Wald, "Sequential Analysis", John Wiley and Sons, N.Y., 1947.

22. A. Dvoretsky, J. Kiefer, and J. Wolfowitz, "Sequential Decision Problems for Processes with Continuous Time Parameter. Testing Hypotheses". National Bureau of Standards Report No. 2023, September 1, 1952.

23. Allen Birnbaum, "The Design of Two Sample Comparative Life Tests", (Paper given at Chicago meeting of Institute of Mathematical Statistics, December, 1952).

24. Leo Goodman, "Calculating Longevity from Sample Composition", report issued by the Statistical Research Center, University of Chicago.

THE TECHNIQUE OF REGRESSION ANALYSIS

B.B. Day, F. R. Del Priore and Edward Sax
U. S. Naval Engineering Experiment Station, Annapolis, Md.

I. Introduction

The previous discussions in this series have been concerned with the study of the measurement of one characteristic, that is, a single variable. However, industrial problems are more often concerned with the study of the relationship between two or more different variables. Typical questions of this nature might be:

If catalyst temperature is raised, will output be improved?
How does speed affect the wear on tires?
Is corrosion a function of concentration of inhibitors?
Is noise correlated with thickness of material?
Is knowledge of thickness of a weld helpful in predicting its strength? Would the prediction be improved with the addition of hardness values?
How do the combinations of oil pressure, oil temperature, speed and load affect output horsepower?

Not infrequently measurements cannot be made directly but must be calculated from a line or curve between two variables; many coefficients and moduli are found in this manner. A precise calibration curve is needed. Theory indicates a certain physical law; experimental work is done to substantiate this theory; how well do the data do this? A prediction formula is needed.

Regardless of the field of application such problems can be handled very nicely by the statistical technique called regression analysis. This technique can answer not only the question as to how well an assumed relationship holds but furnishes the parameters for expressing that relationship in mathematical form.

There has been a great deal of elegant theory developed on this and it still is by no means a finished subject. We will not attempt to go into the theoretical phases of the subject. The list of references will be helpful reading on this. Our purpose is to present the steps in the procedure in regression analysis in such a way that you will be able to apply it on your own problems; to discuss the more general aspects of the subject and finally present a sufficient number of actual examples to indicate the scope of its usefulness.

II. Steps in regression analysis:

a. The example

One small phase of a development program was the determination of the effect of the iron content on the corrosion resistance of 90/10 copper-nickel alloys in sea water. Seven specimens, each containing a particular iron content, were tested in a corrosion-wheel setup. The wheel was immersed in sea water and rotated at 30 ft/sec for 60 days. The corrosion was measured in weight loss in milligrams/sq decimeter/day. The data obtained were as follows:

Iron Content, (%)	.01	.48	.71	.95	1.19	1.44	1.96
Corrosion (MDD)	127.1	124.0	110.1	103.9	86.8	99.2	83.7

How the data were collected, statistically speaking, is an important point. Some discussion of this subject will be given later.

The measurement under study in a regression analysis is called the dependent variable, usually expressed by the letter Y. In this problem, Y would be the measurements of corrosion in weight loss in milligrams/sq decimeter/day. The one or more other variables whose variations may be causing changes in the dependent variable are called independent variables denoted by X or X_i respectively. This problem is the simplest case involving only one independent variable, that is, percent of iron content.

b. Steps in the procedure

Step 1. The first step in determining the influence of iron content on corrosion would be to plot the data, the scattered points in Fig. 1. Such a plot would suffice for some situations.

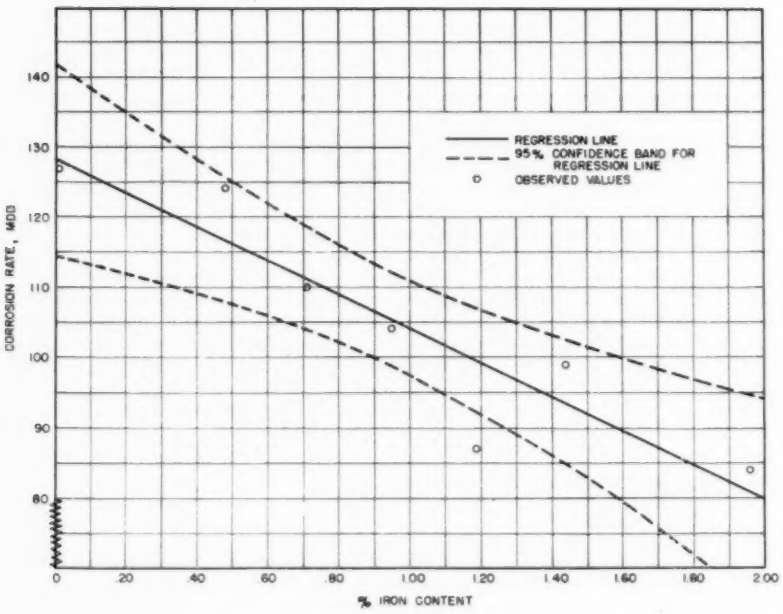


Fig. 1

Corrosion Rate vs Percent Iron Content in 90/10 Copper-Nickel Alloy

However, for this case it is desired to measure quantitatively the effect, if any, of iron content on corrosion.

This plot appears to indicate the simplest type of relationship namely a straight line. The general expression for representing a straight line is $Y = a + bX$. In this problem Y is the weight loss and X the percent iron content, a represents the value of Y when X = 0, the origin of the line, and b, the slope of the line, called the regression coefficient. Thus b is the change in Y for a unit change in X. Here it would be the decrease in weight loss with a unit increase in iron content. In other words, the regression of weight loss on iron content. We shall take as our line $Y_e = a + bX$ where Y_e is the estimated quantity. Then we want values for a and b which will give the best straight line. "Best" is defined as that line which makes the sum of squares of the deviations of observed Y's from the estimated Y's a minimum. Thus we wish $\Sigma(Y-Y_e)^2$ to be the smallest possible. The method for finding this best line is called the Least Squares technique.

Step 2. We will now proceed to find this "best" line. At this stage formulas only will be given, neglecting the details of their development. These can be found in the references.

$$b = \frac{\Sigma(Y-\bar{Y})\ (X-\bar{X})}{\Sigma(X-\bar{X})^2} \qquad \text{.....(1)}$$

$$a = \bar{Y} - b\bar{X} \qquad \text{.....(2)}$$

where $\bar{Y}$ and $\bar{X}$ are the mean values of the measurements of our two variables.

The following sums would be needed to fit the straight line:

$$\Sigma Y,\ \Sigma X,\ \Sigma(Y-\bar{Y})\ (X-\bar{X}),\ \Sigma(X-\bar{X})^2, \text{ and } \Sigma(Y-\bar{Y})^2$$

Using our seven pairs of data: N=7

$$\Sigma Y = 734.3,\ \bar{Y} = 104.90$$
$$\Sigma X = 6.74,\ \bar{X} = 0.9629$$

The short cut methods for the remaining are:

$$\Sigma(Y-\bar{Y})\ (X-\bar{X}) = \Sigma XY - \frac{\Sigma X \Sigma Y}{n}$$

$$\Sigma(X-\bar{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}, \text{ etc.}$$

Thus:

$$\Sigma(Y-\bar{Y})\ (X-\bar{X}) = 127.1 \times .01 + 124.0 \times .48 + \cdots + 83.7 \times 1.96 - \frac{734.3 \times 6.74}{7} = -59.648$$

$$\Sigma(X-\bar{X})^2 = .01^2 + .48^2 + \cdots + 1.96^2 - \frac{(6.74)^2}{7} = 2.4787$$

and

$$\Sigma(Y-\bar{Y})^2 = 127.1^2 + 124.0^2 + \cdots + 83.7^2 - \frac{(734.3)^2}{7} = 1695.19$$

Then substituting in formula (1) above,

$$b = \frac{-59.648}{2.4787} = -24.0642$$

Before using the estimated straight line we should find out how well it fits the data. This will be the next step.

Step 3. The Analysis of Variance technique will be used to test how well the straight line fits the data. The total sum of squares of Y may be partitioned into two parts, the first, with 1 degree of freedom, associated with the regression of Y on X (or the sum of squares due to the regression) and the second, with n-2 degrees of freedom, associated with the variation of Y around the regression line.

Two additional formulas are needed for the sum of squares of the deviations in Y explained by the regression, and for the sum of squares unexplained.

$$SS_{reg} = b\Sigma(X-\bar{X})\ (Y-\bar{Y}) \qquad \text{.....(3)}$$

This becomes $\dfrac{[\Sigma(X-\bar{X})\ (Y-\bar{Y})]^2}{\Sigma(X-\bar{X})^2}$

The unexplained sum of squares becomes

$$\Sigma(Y-Y_e)^2 = \Sigma(Y-\bar{Y})^2 - b\Sigma(Y-\bar{Y})\ (X-\bar{X})$$

Thus we have:

Source of variation	Degrees of freedom	Sum of squares
Explained by regression		$b\Sigma(Y-\bar{Y})\ (X-\bar{X})$
Unexplained	N-2	Difference = $\Sigma(Y-Y_e)^2$
Total	N-1	$\Sigma(Y-\bar{Y})^2$

Carrying through the arithmetic $b\Sigma(Y-\bar{Y})\ (X-\bar{X}) = -24.0638\ (-59.648) = 1435.38$

and $\Sigma(Y-Y_e) = 1695.19 - 1435.36 = 259.83$

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Explained regression	1	1435.38	1435.38	27.62
Unexplained (deviation from regression)	5	259.81	51.96	
Total	6	1695.19		

We can now find the F ratio, explained/unexplained mean square, 27.62, using the usual F tables with which you are already familiar, to find the probability that so large an F would be due to random sampling only.

Here P is less than 01. A small P-value indicates that the straight line accounts for a significant part of the variation in the Y readings. Here it would indicate that a significant part of the variation in weight loss (corrosion) of the seven specimens can be accounted for by the difference in their iron content.

A meaningful value would be the percent of the total variation in the seven readings which is explained by the straight line. This is

$$\frac{\text{SS expl}}{\text{SS total}} = \frac{1435.38}{1695.19} = .847 \text{ or } 84.7\%$$

This value happens to be the square of the correlation coefficient, r. Hence r = .97. However, the value r^2 has the greater significance for it measures the efficiency of the line.

Step 4. Our next step is to determine if we can significantly improve our efficiency by including a quadratic term to our straight line. In other words is the relationship curvilinear? The general expression for this is

$$Y = a + b_1X + b_2X^2 + \text{---}. \qquad \text{.....(4)}$$

In solving a quadratic or cubic, etc., the least squares criterion leads to a set of simultaneous equations. The setting up of these so-called normal equations is quite systematic. First the equation to be fitted is written in the form

$$y = b_1x + b_2x^2 + b_3x^3 + \text{---},$$

the lower case letters here and later indicating deviations from the mean. Since we are fitting the quadratic, two simultaneous equations are needed. The first is obtained by multiplying the equation by the first independent variable (x). The second is obtained by multiplying the equation by the second independent term (x^2). The resultant products in each case are then summed over all the data points giving these equations:

$$b_1\Sigma x^2 + b_2\Sigma x^3 = \Sigma xy \qquad \text{.....(5)}$$

$$b\Sigma x^3 + b_2\Sigma x^4 = \Sigma x^2y \qquad \text{.....(6)}$$

We will need the following additional sums:

$\Sigma(X^2-\bar{X}^2)^2$ that is Σx^4, $\Sigma(X-\bar{X})\ (X^2-\bar{X}^2)$ that is $\Sigma x3$, and

$\Sigma(Y-\bar{Y})\ (X^2-\bar{X}^2)$ that is Σxy^2

The short-cut methods as before are appropriate here. Computing these sums and substituting in the preceding equations (5)and(6) give

$$2.4787b_1 + 4.8913b_2 = -59.648$$

$$4.8913b_1 + 10.6944b_2 = -113.4118$$

Solving these two simultaneous equations give

$b_1 = -32.1918$

$b_2 = 4.1188$

The expression for a would be $a = \bar{Y} - b_1\bar{X} - b_2\bar{X}^2$ but before computing this we will want to determine if there has been a significant improvement in efficiency by adding the quadratic term.

Step 5. To determine this efficiency we again use the Analysis of Variance technique

Source of variation	Degrees of freedom	Sum of squares
Explained (regression)	2	$b_1\Sigma xy + b_2\Sigma x^2y$
Unexplained (dev from reg)	N-3	Difference = $\Sigma(Y-Y_e)^2$
Total	N-1	$\Sigma(Y-\bar{Y})^2 = \Sigma y^2$

Thus $\Sigma(Y-Y_e)^2 = \Sigma y^2 - b_1\Sigma xy - b_2\Sigma x^2y$ .....(7)

(remembering that the x's and y's are in each case deviations from mean).

This becomes	Degrees of freedom	Sum of squares	Mean square
Explained by	2	1453.06	
Unexplained	4	242.13	60.53
Total	6	1695.19	

The improvement due to the quadratic term will be the difference between the sum of squares explained by the quadratic (the above table) and explained by the linear only (Step 3).

	Degrees of freedom	Sum of squares
Explained by quadratic	2	1453.06
Explained by linear	1	1435.38
Improvement with quadratic term	1	17.68

The measure of improvement is then compared with the estimate of random error, here the unexplained by quadratic mean square above, that is 60.53. We do not need to calculate F in this case since the improvement mean square 17.68 is less than the random error 60.53. This means that the quadratic did not give any improvement over the linear. There is no real difference in the estimates of random error by the quadratic and the linear, 60.53 as compared to 51.96. Thus we conclude that the straight line is not only significant but also sufficient. We return now to the results of Step 3 to extend them.

Step 6. We now calculate the a for the linear relationship,

$$a = \bar{Y} - b\bar{X} = 104.97 - (-24,0642(.9629) = 128.14$$

and present the relationship as

$$Y_e = 128.14 - 24.0642X$$

Fig. 1 gives both the data and this line of best fit. The line is drawn between two points which were obtained by substituting values of iron for X in the equation above and finding Y_e, estimated weight loss. Thus, when X = 0, Y_e = a = 128.14 when X = 1, Y_e = 104.08. As a check the line should go through the point $(\bar{X}, \bar{Y})$. The constant value a = 128.14, Y intercept, give the corrosion when iron is not present, b =-24.06 is the slope. The negative sign indicates a decreasing effect. Thus, for every 1% change in iron content, there is a corresponding decrease in corrosion of 24.06 MDD. This measure of the rate of change in corrosion for a unit change in iron content is called the regression coefficient.

Step 7. Since the quadratic term was found to be only random, the unexplained by linear mean square is our best estimate of the random error - variance. In this case it is 51.96. The square root of this is the well-known standard deviation, called s since it is an estimated value.

Hence $s = \sqrt{51.96} = 7.21$

From the analysis of variance form in Step 3 above,

$$s = \sqrt{\frac{\Sigma(Y-Y_e)^2}{n-2}} \qquad \text{.....(8)}$$

This value is also called standard error of estimate. Then the standard error of the regression coefficient, b, will be

$$s_b = \frac{s}{\sqrt{\Sigma(X-\bar{X})^2}} \qquad \text{.....(9)}$$

The formula for the standard error of the regression line will be

$$s_{Y_e} = s\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma(X-\bar{X})^2}} \qquad \text{.....(10)}$$

With these estimates it is possible to put confidence limits about the regression coefficient b, and the regression line itself. Thus for the 95% confidence limits in our example the t-table would be entered with 5 degrees of freedom, and t read from the P = .05 column. This value of t is 2.57

The .95 confidence limits for b would then be

$$b \pm t\, s_b \text{ or } -24.06 \pm 12.28$$

Knowing its standard error, it is possible to compare a regression coefficient with a hypothetical value or with another estimated regression coefficient.

For the regression line confidence bands must be computed for each value of X. The nearer the X is to its mean value the narrower will be the band. To illustrate the computation, consider X = 1.96 substituting in (10) above and multiplying by t gives

$$\pm 2.57 \times 7.21\sqrt{\frac{1}{7} + \frac{(1.96 - .96)^2}{2.4787}} = \pm 13.6$$

Using the estimated regression line

$$Y_e = 128.14 - 24.06 \times 1.96 = 81.0$$

Hence the confidence limits at this value of X are $81.0 \pm 13.6 = 67.4$ and 94.6.

In this section we have covered only the fundamentals of regression analysis. Sufficient references are given for a study of the finer points. We have generally followed the lettering used in reference (4).

III. Extension of the Regression Procedure

a. All the necessary steps for carrying through any regression analysis are given in Part II. To recapitulate we started with the simplest case - a straight line relationship - $Y = a + bX$, showing the steps in computing the best line and testing its significance. From that the quadratic form was introduced and a test made of its adequacy. Similarly, it has been shown how the procedure can be extended to include additional polynomial terms - cubic, quartic, etc.,

$$Y = a + b_1X + b_2X^2 + b_3X^3 + ---,$$

in other words "curvilinear" regression.

b. Finding the relationship of a dependent variable to more than one independent variable is simply a special case of this polynomial form where the other independent variables X_2, X_3 --- appear in the place of the X^2, X^3, etc. The form then becomes

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + ---$$

This is called "multiple" regression.

Thus we have the designations linear, curvilinear and multiple regression with the b's called the regression coefficients. Each b measures the rate of change in Y, the dependent variable, for a unit change in that independent variable.

The procedure for fitting and testing the significance of the regression remains the same for all.

Thus for the multiple case in the normal equations (5) and (6), Section II, x_1 and x_2 would replace x and x^2 and we have:

$$b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 = \Sigma x_1y$$

$$b_1\Sigma x_1x_2 + b_2\Sigma x_2^2 = \Sigma x_2y$$

The normal equations for three variable cases are given in Section V, c.

You will realize that with an increase in the number of terms or independent variables the arithmetic becomes laborious. Schemes for systematizing the computations will be found in the references, (2) and (4) are particularly recommended for this. The scheme we use will be shown in Section V, c.

c. As a special case in fitting a polynomial

$$Y = b_1X + b_2X^2 + b_3X^3 + ---,$$

it is advantageous to have the settings of the independent variable equally spaced and the same number of readings at each X. For such cases as, X = -3, 0, 3, 6, etc., there are tables available which lead to a rapid solution even of polynomials of high order. References (3) and (9). Using these tables of coefficients called ϵ', each term can be fitted by the simple formula

$$b = \frac{\Sigma Y \epsilon'_i}{\Sigma \epsilon'^2} \qquad \text{(i denotes the order)}$$

The steps in the arithmetic are outlined in detail in reference (8). This is similar to the linear formula (equation (1)) except that by reading the corresponding ϵ' in place of our x values in these tables, the same simple formula applies even if the b is for the second degree term or third degree term, etc. In addition and oftentimes very important the b's for the different terms are independent of each other. This is the basis for calling these the "orthogonal polynomial" tables. Testing the significance of each additional term in the polynomial is the same as for the linear case, Section V, b.

For test work that can be carried on with equally-spaced intervals of the independent variable, this technique is invaluable. An example of its use will be described later.

d. Covariance. We have seen how the analysis of variance technique is an integral part of regression analysis. Conversely regression techniques can be an important adjunct to the analysis of variance. This entire development is due to RA Fisher (2).

Such variable conditions as these mentioned may have significant effect on the measurement under study, thereby obscuring the true results of the test.

Due to physical limitations it may not be practicable or possible to keep certain variables (not under test) in control. For example in comparing weathering resistance of different paints, the amount of paint applied may vary. Or it would not pay to build a constant temperature room for studying thermal use in electric motors. Planned Experiments on a production line must usually be run without having all variables under control. Original specimens used in testing may vary from specimen to specimen in size, weight, etc.

An extension of regression called the Analysis of Covariance is a means of equalizing the influence of these variables or uncontrolled conditions provided they are measurable. References (2), (5) and (7).

At the same time the influence they have had on the effect under study can be evaluated numerically. This technique adjusts the analysis of variance on the effect (Y) for the variation in a variable (X) using regression procedures. In fact, the expression $\Sigma(Y-\bar{Y})(X-\bar{X})$ or Σxy in II above, is called the covariance.

The procedure requires that the X values be analyzed in exactly the same way as were the Y values, the measurement under study. A third setup with exactly the same analysis of variance breakdowns will be that of the covariance of Y and X, reference (2) article 49.1.

The procedure will be demonstrated in an actual situation in the last problem, Section V. It illustrates the linear case. Curvilinear and multiple would just be an extension of the procedure. This problem is the kind that demonstrates both the equalizing influence and the evaluation of the relationship.

This technique makes it possible to discriminate as to whether the influence of an independent variable is quantitative or qualitative. For example, is the effect of more than one coat of a rust arrestor due to the amount of compound applied or to the drying out period between coats.

IV. Discussion

a. The selection of the dependent variable in a regression analysis is usually by a "cause" and "effect" relationship. The independent variable would be the "cause" and the dependent would be the "effect". However, the independent variable may not necessarily be a causal variable. For example, the independent variable may be many variables removed from the direct cause. The results would then be interpreted with this in mind. Perhaps no cause and effect relationship is even hypothesized. Then the criterion may become the use that is expected to be made of the analysis, etc. However, in these situations the difficulty in selection of independent and dependent variables causes us to be wary of our interpretation of the results.

b. The settings or values of the independent variable may be either arbitrarily or randomly selected. There may be one measurement of Y for each X value, an equal number, or a variable number of Y's for each X value.

c. A very important assumption in the formulas used is that independent values are known to be free from error. This can be relaxed to relatively free from error as compared to the error in the dependent data which are the subject of study. It is important to note that the deviations from the line in the example, Section II, are vertical distances, that is, deviations of Y value. This follows from the assumption that only Y is in error, X is relatively free from error. If the opposite situation existed, the Least Squares criterion would give a slightly different line since we would be minimizing horizontal distances. The formulas would be similar except for interchanging X and Y. Thus there are two regression lines possible, Y on X and X on Y. Between these two extremes, vertical and horizontal, there are various approaches, depending upon our knowledge of the errors in both X and Y variables.

d. Frequently there is another measure of the random error involved in the dependent variable besides the error from regression. This usually comes about from running replicates, etc. at each value of the independent variable. Then, comparison of these two measures of error would be very informative.

Using our example as an illustration, if two concentrations had been made, i.e., from different mixes, there would be a measure of error independent of the regression. If the mean square from regression were significantly larger than this experimental error, the inference would be that the relationship can still be improved - perhaps with a different type of curve. Should the experiment error happen to be significantly larger than regression error, then it would certainly be time to start looking at the details of the test setup.

It is worthwhile to call attention to the fact that this experimental error is not derived from pairs of readings on the same specimen, say duplicates. The duplicate error would only be a measure of the reading error and as such is not an estimate of a true experimental error.

Along this vein it is vital that randomness enter into the testing at many stages. This is, of course, necessary to prevent any systematic effect from distorting the results. For example, suppose in our problem, the heat treatment was given to the various slabs in an orderly fashion. The material containing the lowest iron content was heat-treated first, second lowest treated second, etc. A question would then arise as to how much of the effect that we were attributing to iron content was actually due to temperature regulation in our heat treatments.

e. The question may be asked, what about correlation? Little has been said about the correlation coefficient. This omission has been intentional because in the great majority of problems other statistical constants are more meaningful. These have been mentioned in Section II above. One such is the square of the correlation coefficient,

$$r^2 = \frac{\text{SS in Y explained by regression}}{\text{Total sum of squares of Y values}}$$

In fact, in practice we actually use a probability table for r^2 computed from the r table. Also the b's, the regression coefficients, are extremely useful since they measure the rate of change in the dependent variable with unit changes in the independent. They are either positive or negative thereby indicating whether the change is an increase or decrease.

V. Problems in Regression Analysis - A few actual examples of the uses made of regression may be helpful.

a. A calibration problem in linear regression. The Chemical Engineering Laboratory of the Station was interested in the use of the infrared spectograph in place of the time-consuming chemical methods in the determination of additive content in Diesel engine oils. A calibration curve of concentration and density was needed for the particular wave-length to be used. It was important to know the error in the calibration curve for comparative purposes.

There were two interesting aspects to this example which will be brought out in the discussion.

The experimental data consisted of from two to eight readings of density (Y) at each of eight concentrations (X), making a total of 37 items. A straight line was found to be adequate. It was

Density = .00465 + .0147 Concentration

The corresponding analysis of variance was:

Source of variance	D/f	Mean Square	F	
Explained by linear	1	0.720033	102.2	Highly significant
Deviation	6	.00001285	2.2	Not significant
Between concentrations	7			
Within concentrations	29	.00005844		
Total	36			

That the "deviation" from the straight line was not significantly larger than "within concentrations" indicated that the straight line was an adequate mathematical expression of the relationship. This verified the accepted theory that it should be a straight line (Beer's Law). It is important to point out that the "within concentrations" were from density readings at the same percentage concentration, but on different solutions, i.e., made up on different days.

Beer's Law was therefore assumed to hold in calibrating the instrument again after its periodic cleaning. This assumption led to only two concentrations being tested; the lowest and the highest percentage in the zone in which we were interested. As reference (6) points out, these are the most efficient settings in determining the straight line. The saving in testing to determine the straight line permitted an incorporation of a study of other factors: preparation, stability, variable drive. The method of covariance was used in this study to take into account input voltages and operating temperatures.

Consider the calibration line we obtained which was to be used in predicting the concentration from the density readings. Since we are predicting the independent variable, we are interested in the error of this prediction, a switch-about from the usual situation. This required a refinement in the error formula, reference (1), page 133 gives the variance of this prediction as:

$$\text{Variance } (X) = \frac{s^2}{b^2}\left\{\frac{1}{m} + \frac{1}{n} + \frac{(Y-\bar{Y})^2}{b^2\Sigma(X-\bar{X})^2}\right\} \qquad \text{.....(11)}$$

where Y stands for density, X for concentration, and m the number of readings at any particular point.

b. Fitting by Orthogonal Polynomials. Relationship of output frequency to load in the Vickers 10 KN Motor-Generator Set.

This generator set was a new design sent to the Electrical Laboratory for testing to specification requirements. Because the project engineer was interested in studying its overall characteristics, statistical planning was used. The analysis of variance indicated that load, a main effect, was a significant factor on output frequency. It was desired therefore, to find out if there was an evident trend in output frequency with load. Even though an analysis of variance fails to show significance due to a variable, it is still important to look for trends. Regression analysis is the technique for doing this.

Load (X)	0	1/4	1/2	3/4	1	1-1/4
Output freq (Y)	399.83	399.58	399.36	399.20	398.84	397.98

Here load will be the independent variable, output frequency the dependent. The data plotted in Figure 2 indicated that the magnitude of output frequency was associated with load changes. Since the load settings had been selected at equal intervals, the regression analysis was done by the use of orthogonal polynomial, references (3) and (8).

The table below gives the analysis of variance test up to and including the cubic. It was sufficient since the residual mean square .0021 was of the same order of magnitude as was the experimental error variance of test.

Analysis of Variance

Source of variation	D/f	Sum of squares	Mean square	F	Probability levels .05	.01
	(1)	(2)	(3)	(4)	(5)	(6)
Total	5	2.1569	-	-	-	-
Degree 1	1	1.9322	1.9322	34.38	7.71	21.20
Residual 1	4	.2247	.0562	-	-	-
Degree 2	1	.1551	.1551	6.69	10.13	34.12
Residual 2	3	.0696	.0232	-	-	-
Degree 3	1	.0654	.0654	31.14	18.51	98.49
Residual 3	2	.0042	.0021	-	-	-

The fitted line has been drawn in Figure 2. Thus, there was a somewhat rapid decrease in output frequency with the addition of load, then a slackening in this decrease but followed at the higher loads with an accelerated change. It should be emphasized that this example demonstrates two important points:

(1) Use of orthogonal polynomials when the X-values are equally spaced.
(2) Regression analysis included in the analysis of variance will give a more precise interpretation of the main effects.

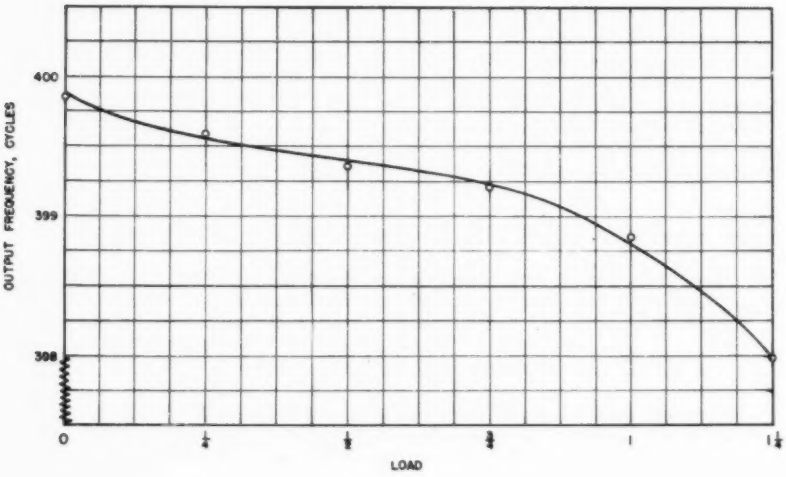


Fig. 2

Output Frequency vs Load on a Motor-Generator Set Fitted by a Third-Degree Polynomial in Load

c. Multiple regression in a bearing research problem. One phase of a test program on the study of bearing materials was concerned with evaluation of the influence of varying factors on bearing performance. One set of factors selected was speed, load and the viscosity of the lubricating oil. Performance was measured in the friction generated. Using a 60° segment bearing the test was run on the standard Amsler abrasion testing machine adapted to make use of basic frictional measuring devices under varying speeds and loads. A 7 x 7 lattice plan allowed for 49 different speed-load combinations, viscosity being measured at each of these.

The theoretical functional relationship based on engineering judgment was

$$f = b_o N^{b_1} P^{b_2} Z^{b_3}$$

where N = speed, P = load, Z = viscosity and f = friction; the b's, the parameters dependent on test setup, to be estimated from the experimental data. The above expression is in line with Hersey's well-known $\frac{ZN}{P}$ factor.

This problem illustrates the situation where a transformation is necessary to give an equation of the linear form in order to use the Least Squares criterion. Using the logarithmic transformation the new equation was

$$\log f = \log b_o + b_1 \log N + b_2 \log P + b_3 \log Z$$

Placing:

$$Y = \log f,\ a = \log b_o,\ X_1 = \log N,\ X_2 = \log P \text{ and } X_3 = \log Z.$$

this may then be expressed in the standard regression form

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

The corresponding normal equations were then

$$b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 + b_3\Sigma x_1x_3 = \Sigma yx_1$$

$$b_1\Sigma x_1x_2 + b_2\Sigma x_2^2 + b_3\Sigma x_2x_3 = \Sigma yx_2$$

$$b_1\Sigma x_1x_3 + b_2\Sigma x_2x_3 + b_3\Sigma x_3^2 = \Sigma yx_3$$

The first step in the arithmetic was to change each of the 49 sets of readings of speed, load, viscosity and friction to logarithms. The next step was then to compute the required sums of squares and cross products of these 49 sets and correct them in each instance to the mean values. This is shown in Part 1 of the work sheets reproduced here. These sheets give the entire layout of the procedure we follow in multiple regression problems.

Briefly, instead of solving the three simultaneous equations directly for the b's, what is known as the Gauss multipliers were used, those are c_{11}, c_{12}, --- c_{23} --- c_{33}, reference (4), page 139. Using the same summations as in the left part of the normal equations and these c's, three sets of equations may then be written with the right members being 1, 0, 0, and 0, 1, 0 and 0, 0, 1 respectively. These are indicated in a compact form in Part 2 of the work sheets. The solution of these yield a matrix of the c values, that is

$$\begin{matrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{23} & c_{33} \end{matrix} \qquad \text{where } c_{21} = c_{12} \text{ etc.}$$

There are definite advantages in using the c - multipliers. Frequently it happens that, for the same set of independent factors, it is desirable to examine the regressions for more than one dependent factor. Using the c's, the simultaneous equations will only need to be solved once. They greatly facilitate the computation of the standard errors of the regression coefficients. And finally they provide a method for readily finding the new b's when a variate is omitted, reference (2), article 29.1.

Part 2 of the work sheets shows the method we used in solving the normal equations. Any other method might have been used, a number of which will be found in the references.

The analysis of variance for testing the regression is shown in Part 2 of the work sheets. The relationship was significant accounting for 71% of the variation in the log friction values.

WORK SHEETS FOR MULTIPLE REGRESSION - PROBLEM c

Part 1. Correcting sums of squares and cross products to their means.

N = 49	Speed X_1 = log N	Load X_2 = log P	Viscosity X_3 = log Z	Friction Y = log f
Sum	110.523	111.091	95.125	112.643
Mean	2.256	2.267	1.941	2.299
Speed, X_1	252.590347	253.012984	212.474692	254.011640
Correction factor	249.292521	250.573686	214.561232	254.074332
Difference	3.297826	2.439298	-2.086540	-0.062692
Load, X_2		260.325115	212.630421	252.029890
Correction factor		251.861434	215.663906	255.380072
Difference		8.463681	-3.033485	-3.350182
Viscosity, X_3	This half a mirror view of above.		186.629107	219.588080
Correction factor			184.668686	218.676844
Difference			1.960421	0.911236
Friction, Y				261.329821
Correction factor				258.947866
Difference				2.381955

Part 2. Solution of the normal equations using c multiplier, reference (4).

	3.297826	2.439298	-2.086540	1,0,0,	...(1)
		8.463681	-3.033485	0,1,0,	...(2)
A mirror of that above line.			1.960421	0,0,1,	...(3)
(1)1.960...-(3)(-2.086...)	2.111471	-1.547436	= 1.960420	8, 0,	...(4)
(2)1.960...-(3)(-3.033...)		7.390346	= 0	, 1.960420 8,	...(5)
(4)7.390...-(5)(-1.547...)	13.2099885		= 14.4881886		...(6)

c_{11} = 1.096760 c_{21} = 0.229646 c_{31} = 1.522664
c_{22} = 0.313353 c_{32} = 0.729290
c_{33} = 3.259194

The b coefficients are

b_1 = 1.09676(-0.062692)+0.229646(-3.35018)+1.52266(0.911236) b_1 = 0.54939
b_2 = 0.229646(-0.062692)+0.313353(-3.35018)+0.729290(0.911236) b_2 = -.39963
b_3 = 1.52266(-0.062692)+0.729290(-3.35018)+3.25919(0.911236) b_3 = 0.43118

For the analysis of variance

SS_{reg} = 0.54939(-0.062692)+(0.39963)(-3.35018)+0.43118(0.911236) = 1.697264

Source of Variation	D/f	SS	MS	F
Regression	3	1.697264	0.56575	37.18
Unexplained (error)	45	0.684691	0.0152154	
Total	48	2.381955		

$R^2 = \frac{1.69726}{2.38196}$ = 0.71 or 71%

The standard errors are:

s = $\sqrt{0.0152154}$ = 0.1234

s_{b_1} = 0.1234 $\sqrt{1.09676}$ = 0.1292

s_{b_2} = 0.1234 $\sqrt{0.313353}$ = 0.06909

s_{b_3} = 0.1234 $\sqrt{3.259194}$ = 0.2229

We will now consider the effect of the three factors, alone and in combination with the other factors, using as our criterion, R^2, the percent of the total variation in the log frictions accounted for by the regression.

One factor	R^2 percent	Two factors	R^2 percent
Speed	<1	Speed and load	69
Load	56	Speed and visc	49
Viscosity	18	Load and visc	60

Thus while speed alone indicated no significant effect on friction, $R^2<1\%$, in combination with load there was a considerable improvement over load alone, 69% against 56%. This illustrates a very important point, namely that the test of a factor alone is not sufficient to indicate its significant influence. There can be an accelerated effect in combination with another factor.

Viscosity in combination with speed and load has little or no additional influence on friction. It seems that there is an area of over-lapping effect of speed and viscosity, that is a tie-in between the two. This is a reasonable conclusion since it is known that viscosity changes with speed.

As to be expected, the b_2 value (load) is negative. This would put it in the denominator in the original function, again in line with the $\frac{ZN}{P}$ concept.

The numerical value of the b's and their standard errors are found in Part 2 of the work sheets.

This and other studies indicated that the Hersey variable $(\frac{ZN}{P})$ works well with the exponent of around 0.5. This was attributed to the use of segment bearings.

d. A problem applying the analysis of covariance. The technique of the analysis of covariance is illustrated in one phase of a rust arrestor problem. Steel coupons, already in a measurable rusted state, were coated with a rust arrestor compound by one of three methods: brushing, spraying, and dipping. Five coupons coated by each method were then exposed in the salt spray cabinet. The additional rust due to this exposure was measured. This was the effect (Y) under study; the analysis of variance of the measured results was as follows:

Source of variation	D/f	SS	MS	F
Methods	2	520	260	3.5
Error	12	884	73.67	
Total	14	1404		

The (F here) was too small to indicate a definite significance between methods.

The condition of the coupons prior to coating, which could not be controlled, was then brought into the picture by the analysis of

covariance. This initial rust (X) was found to be highly related to the rust due to exposure. In addition, when the coupons were all put on an equal footing with regard to this initial rust, the test became more sensitive due to the reduction in the experimental error. Differences between methods became highly significant. The procedure for detecting this follows. The corresponding weight change (grams coded) measuring the rust due to exposure (Y) and the initial rust (X) for the five coupons at each method were as follows:

	Brushing		Spraying		Dipping	
	Y	X	Y	X	Y	X
	63	16	81	48	72	40
	77	45	73	40	54	31
	81	50	59	24	57	40
	60	19	74	33	59	33
	63	18	77	41	52	20
Total	344	148	364	186	294	164

As stated in Section III, the sum of squares for the three components Y, X, and their covariance YX were put into an analysis of variance table following the standard procedure for computing the breakdowns. These are given in columns 2, 3, 4 in table below.

The Analysis of Covariance

Source of variation	D/f	$\Sigma(Y-\bar{Y})^2$	$\Sigma(Y-\bar{Y})(X-\bar{X})$	$\Sigma(X-\bar{X})^2$	$b=\frac{(3)}{(4)}$	Explained by regression $(3)\cdot(5)$
	(1)	(2)	(3)	(4)	(5)	(6)
Methods	2	520	100	145		
Error	12	884	1075	1687	.6372	685.0
Total	14	1404	1175	1832	.6414	754

Source of variation	Adjusted by regression SS_y (2)-(6)	D/f	MS	F
	(7)	(8)	(9)	(10)
Methods	451*	2	225.7	12.5
Error	199	11	18.09	
Total	650	13		

* By subtraction $SS_T - SS_E$

The adjusted sum of squares for error, column 7, was found to be 199. This represented the deviations within methods from an average within-methods regression. The adjusted mean square was 18.09. This was a much more sensitive error term than the unadjusted error estimate of 73.67, it shows the importance of initial rust on rust due to exposure. The adjusted sum of squares for the total, column 7, was found to be 650. This represented the deviations from a regression line based on all 15 points. The difference 650 - 199 = 451 was the sum of squares associated with differences among methods.

The F value (12.5) was highly significant; thus, we conclude that methods were different when the rust due to exposure was adjusted for the initial rust.

VI. References

1. Owen L. Davies, Statistical Methods in Research and Production, Chap. 6, Oliver and Boyd.
2. R. A. Fisher, Statistical Methods for Research Workers, Chap. V, Arts 25 - 29.2, Chap. VI, Chap. VII, Arts 44 - 47, 49.1, Oliver and Boyd.
3. R. A. Fisher and F. Yates, Statistical Tables, 3rd Edition, Oliver and Boyd.
4. C. H. Goulden, Methods of Statistical Analysis, 2nd Edition, Chaps. 6 - 9, John Wiley & Co.
5. G. W. Snedecor, Statistical Methods, 4th Edition, Chap. 6, 12, 13, 14, The Iowa State Press.
6. Cuthbert Daniels, Design of Experiments for Most Precise Slope Estimation, Industrial and Engineering Chemistry, Vol. 43, No. 6, June 1951 - page 1298.
7. B. Day and R. A. Fisher, The Comparison of Variability in Populations Having Unequal Means. An Example of the Analysis of Covariance --- Annals of Eugenics, Vol. VII, Part IV, 1937.
8. F. R. Del Priore and B. Day, An Application of a Simple Technique in Curve Fitting of Certain Types of Data, The Instrument Maker, Sept - Oct 1948, V. 16, #5.
9. D. B. DeLury, Values and Integrals of the Orthonal Polynomials up to n = 26, University of Toranto Press, 1950.
10. Lois M. Zucker, Evaluation of Slopes and Intercepts of Straight Lines, Human Biology, December 1947.

VII. Acknowledgements are made to the technical staff of the U. S. Naval Engineering Experiment Station in permitting use of experimental tests as examples in this paper.

The opinions and assertions in this paper are the authors' and are not to be construed as official or reflecting the view of the Navy Department or the Naval Service at large.

VENDOR RESPONSE TO THE QUALITY CONTROL APPROACH

C. R. Burdick
Ford Motor Company

Good Afternoon, Gentlemen:

I welcome the opportunity to take part in this session and to contribute to the subject, "Vendor Response To The Quality Control Approach."

I realized, when I was first invited to talk on this subject, that all of my remarks would be made from the point of view of a consumer ... a discussion, mainly, of what we at the Ford Motor Company had experienced in our Vendor Contact Program. That, to me, had the appearance of being a one-sided story. In my mind, what the suppliers thought of our approach, and their general attitude to Quality Control, was just as important as what we thought of their response. So I took the liberty of inviting representatives from two of the Ford Motor Company's suppliers to be here with me today to present their own views in their own words. These gentlemen are:

Mr. Donald Spath, Chief Inspector, General Products Corp., Jackson, Michigan

Mr. Kenneth Rosengren, Manager, Quality Control Houdaille-Hershey Corp., Houde Engineering Div., Buffalo, New York

I am sure you will be particularly interested in hearing from them later in this session.

First, however, I think it would be a good idea to acquaint you with the historical background of the Ford Motor Company's Vendor Contact Program ... how the idea originated ... the formulation of our plan ... and its practical administration.

As might naturally be expected, our desire to exchange ideas with our suppliers on modern quality control techniques stemmed from the success we had enjoyed in applying these techniques to our own operations. We had adequate proof that Quality Control meant Quality Improvement. It was a fundamental, however, that we could not substantially improve the quality of our end products by improving only that percentage of parts and material which we ourselves produced. We had to have the support and cooperation of our suppliers. Thereby was born the idea for a Vendor Contact Program.

But let me impress upon you now that our motives for engaging in such a program were by no means selfish. While we were very definitely looking out for ourselves insofar as we were determined to improve the quality of our own products, we were also convinced that our suppliers would also benefit from the adoption of statistical quality control techniques. We reasoned that if we, at the Ford Motor Company could reduce our rejects and repairs by applying these new methods, so could our suppliers! If we could cut down on our costs, so could our suppliers! If we could instil better human relations, so could our suppliers!

Viewed from any angle, it was evident that statistical quality control was good for both the consumer and the supplier.

Convinced of the mutual benefits to be enjoyed, we in Quality Control presented our case to our Purchasing people. We explained the idea of a Vendor Contact Program and substantiated all of our claims with actual examples of quality improvement and cost reduction in our own operations. The Purchasing people were impressed, but not sold. The query in their minds was: "What do other companies think about Quality Control"?

How better could that question be answered than by a visit to a Quality Control convention? And so it transpired that one of our Purchasing Agents visited a convention such as this in Boston.

Despite what we had told him about statistical sampling, he proceeded with a one hundred percent cross-examination.

I think he talked to practically everyone there ... question after question ... in an effort to determine what industry, in general, had to say about Quality Control.

He came back a fellow enthusiast! Shortly thereafter our Vendor Contact Program was officially initiated.

Laudable as we knew our intentions to be, we considered our plan of approach to be of extreme importance. We did not, for instance, wish to give the erroneous impression to any of our suppliers that we were trying to foist something upon them. Just as we consider ourselves to be specialists in certain fields, so do we acknowledge that our suppliers have accummulated specialized knowledge in their fields. They know more about their business than we do and we had no intention of trying to tell them how to run it. We did, however, want to tell them about statistical quality control and of its far-reaching advantages; indeed, it was our genuine belief that we would be performing a dis-service if we failed to do so. Conversely, we felt that the best service we could perform would be to offer some concrete assistance in the explanation and application of the technique. So it was decided that training courses would be made available at the Ford Motor Company Training Department, and that personal assistance would be given, at the suppliers' plants, by our own people, in the installation of quality control techniques.

The official approach was made early in January, 1950, when a letter under the joint signature of the Vice President - Purchasing, and the Director of Quality Control, was mailed to some 600 of our major suppliers. Enclosed with this letter was a small pamphlet - "Don't Gamble With Quality" - which we had developed for the information of our own supervision. This pamphlet described the basic principles of statistical quality control and was intended to give our suppliers a quick insight into the technique. Also mailed was a brief questionnaire on which our suppliers could indicate if they wished to have a representative call upon them for further discussion or if they wanted to avail themselves of the training sessions offered.

Now came the full significance of the question: "What would the vendors' response be"? Would they think us presumptuous? Would they

reject the idea or welcome it? There was nothing to do but wait and see.

The results surpassed our expectations. As we had hoped, American industry had again demonstrated that unique American ability to compete and cooperate at one and the same time. One hundred and seven of our suppliers sent key men from their organizations to the training sessions at the Ford Motor Company. In all, 258 men attended and received a one and one half day introductory course in statistical quality control, which included a tour of our plants to see the technique in practical application.

This training program was repeated in the fall of the same year with an attendance of 228 men from 85 of our suppliers, and on the West Coast a similar program was conducted for 101 representatives from 20 West Coast firms. In addition, a total of 164 firms were visited, upon request, by Ford Motor Company Quality Control personnel. This, in itself, involved over 37,000 miles of travel in 21 States.

What would the vendor's response be? For us, that question was still only partially answered. To be sure, we were gratified by the initial response to our idea, but the proof of the pudding as far as we were concerned would be the application of statistical quality control by our suppliers and the resulting improvement in Quality of their products.

What better proof than Quality Level Certification? That is the most advanced stage of "Vendor Response." To date, we have 301 parts covered by Quality Level Certification. It is but a modest beginning, but one of which we are proud, and which leads us to the conviction that the day is not far distant when Certification will be the rule rather than the exception.

It would be wrong for anyone to assume, however, or for me to give the impression, that Quality Level Certification is an easily achieved status. It is not. On the part of the consumer and supplier alike it requires something akin to a passion for quality improvement plus the ability to satisfy that desire. But as is usual in such cases, the reward outstrips the effort.

A problem that industry has always faced has been the lack of coordination between supplier and consumer quality programs. In the past a supplier was given the blueprints and specifications for the part he had contracted to make. He was told the quantity required and the expected dates of delivery. In regard to the manufacture of the part, however, much was left to his own discretion. The suppliers did a remarkably fine job. But inevitably, there were occasions when shipments inspected by the consumer were found to be unsatisfactory.

In many cases, the supplier's failure to meet quality requirements was due to a lack of understanding rather than a lack of good intention. In some instances, the supplier did not know which dimensions were more important than others, with the result that critical details were sometimes overlooked while relatively unimportant details received concentrated attention. Then, there were cases where the supplier and the consumer used different inspection equipment and different inspection methods. Naturally, they came up with different

answers. The supplier's inspection indicated that he was meeting quality requirements. The consumer's inspection indicated that the shipment was unacceptable. Result: strained relations.

Quality Level Certification is tailor-made to overcome these limitations. In a nutshell, certification amounts to the dovetailing of supplier and consumer quality control programs to insure mutual understanding of all factors affecting quality. In effect, a written agreement is entered into in which both the supplier and the consumer agree upon the level of quality to be maintained and the means by which conformance to quality standards will be assured and measured.

There are seven basic steps to Quality Level Certification. First, blueprint and material specifications must be thoroughly reviewed to determine their relative importance. Second, the specifications should be classified into basic groups, descriptive of their degree of importance. At the Ford Motor Company, we have established four such groups: critical, separate, major and minor. In the critical group are classified those characteristics which vitally affect safety, and which at all times must completely adhere to specifications. Separate characteristics are chemical, metallurgical or other like properties, which because of the restrictive laboratory methods required to determine conformance, can not be economically subjected to statistical sampling. Here, again, complete adherence is required. Major characteristics are those which significantly affect the function, effectiveness, utility, life and appearance of the product, and minor characteristics are those which require general adherence to specifications. Once the characteristics have been classified, tentative quality levels ... which the supplier is required to meet in order to make the shipments acceptable to the consumer ... are established for the majors and minors. As a result of this action, both consumer and supplier have a mutual understanding of the quality standards to be maintained.

The third step involves the correlation of inspection equipment and methods. To do this, the consumer and supplier jointly determine which gaging techniques should be used, how much inspection, and what type of inspection, should be performed to assure conformance to specifications. To avoid any possibility of differences in inspection results, and to afford the maximum assurance that quality standards are being met, statistical sampling tables are used. These tables show how many parts should be inspected in a lot of a given size. Acceptance or rejection is then based on the number of defects found in the sample.

The fourth step in the preparation for Quality Level Certification involves a period of trial during which records are kept by both the supplier and the consumer to measure quality performance.

When the results of this trial period indicate that consumer and supplier quality control programs are closely integrated, final agreement is reached as to the levels of quality which the supplier must attain to assure acceptance by the consumer. This is the fifth step toward Quality Level Certification.

The sixth step is the formalizing of the Agreement. This is a document upon which quality characteristics are listed in groups in

order of importance, and specifying the quality levels which the supplier has agreed to maintain and which the consumer has agreed are acceptable. This becomes an attachment to the Purchase Order. The final step is the actual administration of the Agreement.

The advantages of such a program of Quality Level Certification must surely be obvious. In the first place, both the consumer and the supplier benefit from improved quality in their own operations. This improvement in quality, and the resultant decrease in scrap and rework, means a reduction in cost. And finally, but by no means least, come the tremendous advantage of sound, pleasant business relations. In fact, at the Ford Motor Company we do not think in terms of consumers and suppliers. We think in terms of partners.

Gentlemen, I think that should answer your question: "What was the vendors' response"?

Now, I would like the representatives from two of our suppliers to tell you their own story. First, let me introduce Mr. Donald Spath, Chief Inspector, General Products Corp., Jackson, Michigan.

Good Afternoon, Gentlemen:

In my opinion, Mr. Burdick has very ably covered the subject, "Vendor Response to the Quality Control Approach."

As he has pointed out, his remarks have stemmed chiefly from the point of view of a consumer. Right at the outset, then ... and as one of the business partners of which he spoke ... I would like to substantiate his statement that supplier as well as consumer benefits greatly from a joint, genuine endeavor to improve quality. This has certainly been the experience of the General Products Corporation.

Relatively speaking, our company is a small one, employing at this time about 170 men. We manufacture automotive parts, and our manufacturing operations consist of turning, boring, facing grinding, milling, drilling, threading, broaching, plating, painting, and heat treating on both castings and forgings.

We started our quality control program in June, 1949, at the suggestion of, and with the help of The Ford Motor Company. It was their suggestion to pick one job and stick with it until it was running "in control." We selected a facing operation on a gear shift lever arm that had been very troublesome to us. We had experienced considerable down-time and rejections at final inspection ran as high as 8 to 10%.

Before starting quality control on this job, we explained to the inspectors, operators, and foremen what our plans were and how we proposed to operate. There was the usual skepticism of a few people, but as good results began to appear, this attitude soon changed.

After charting this job for a couple of weeks, we found that changes had to be made in order to meet specification limits and quality control limits. We had to reinforce the machine spindle, strengthen the work-holding fixture, and make a few minor adjustments,

as well as a slight change in the tool grind. These changes, plus the fine cooperation of all concerned on this job resulted in a reduction in rejects, increased production, and improved morale of the operators. Rejects on this operation were reduced to less than 1% and the slotting operation which followed was also improved due to reducing the face runout from as high as .005 to about .001.

The success achieved on this job proved to operators and supervision alike the merits of quality control, and we gradually expanded our program to cover our machining and heat-treating operations.

As Mr. Burdick pointed out, we, as a supplier, have realized the same advantages as the consumer: reductions of scrap and rework, resulting in savings to the company; reduced maintenance cost due to the prompt correction of faulty machines; better cooperation between operators, supervision, and inspection; and improved customer relations due in some measure of course, to improved quality of product.

Conversely, the benefits of our quality control program are being realized by our customers. The quality of our product has improved to such an extent that we are able to certify the quality characteristics of our product and one large customer has been able to reduce receiving inspection costs on our parts 80%.

I hope the story of our experience may be a source of encouragement to other small companies who are contemplating the use of quality control in their own plants.

Thank You

Now I would like to have the Quality Control Manager of one of our larger suppliers tell you how he and his company feel about our approach and the Vendor's Response to the Quality problem ... Gentlemen, Mr. Kenneth Rosengren, Manager, Quality Control, Houdaille-Hershey Corporation, Houde Engineering Division, Buffalo, New York, who will speak for the entire Houdaille Corporation.

Good Afternoon, Gentlemen:

In discussing the matter of Quality Control it has been our thought that to get out of it what we would like we must put into it what is necessary.

In this instance we wish to discuss our relationship on this subject with Mr. Burdick's company, the Ford Motor Company.

Early in our Quality Control History we were approached by the Ford Motor Co. to give consideration to a further expansion of our Quality Control setup, as well as giving thought to presenting our merchandise on a certified basis. We were not only happy to give their suggestions consideration but very pleased that we were going to be able to discuss our problems inspection-wise on a common level, where our tools would be common to both parties. This probably is the prime factor in our acceptance of the overall Quality Program. Believe me our work was cut out for us. It was necessary to train employees to properly deliver to us the results we set out to

accomplish as well as to sell a new tool to our production. In the main this was not hard to do because the time was spent to properly acquaint our supervisory group with our plans. We also took the time to point out advantages that could accrue with this new approach.

Like many another new approach to a problem we had difficulties at the start. Our aims were high but our immediate results fell short of our goal. This goal being certification of our products to our customer.

We decided to make our door lock our first product to be sold on a certification basis. This lock incidently posed quite a problem at the time Quality Control methods were applied. In gathering our original data, many items that went into the final assembly were found to be somewhat beyond a good Quality Level on a consistant basis. Our immediate approach was to find our worst spot and camp on it until some semblance of consistancy was obtained. This policy was carried out on all parts manufactured by us until all parts made at our plant were covered.

Simultaniously with this action we investigated our purchased parts and found our acceptance methods were not as good as could be desired.

Our approach to this problem was a sampling plan that would give us the assurance needed quality-wise. There were instances where vendors quickly made any necessary changes to bring their quality level up to par and others who resisted our efforts to do so. In the latter case, we found it advisable to have the vendor visit our plant to see first hand the problem created by poor quality material. At the time of the visit Quality Control was discussed, with the thought in mind to convert our vendor to this type of operation which would assure him better quality.

Very accurate records were kept on our vendors. These records told a progress story of a most fascinating nature. In most instances the supplier improved his quality to a point desired but a few fell by the wayside and had to be replaced by others.

After our original attempts at a Quality Control application we found our scrap and rework excessive.

This was watched carefully with results that to the neophyte were startling to say the least. To give you a more concrete answer to what we found, that over $80,000 was saved in scrap and rework alone in the first six months. This is only part of the savings. We gradually reduced our inspection force and scrap and rework were reduced as noted. These things of a tangible nature make for good conversation but there are intangibles quite as important. Perhaps the most important of the intangibles is selling a product on a certified basis. In the instance of the door lock, when we felt we were ready to certify our lock production we invited representatives of our customer to visit our plant and go over our Quality Control Program, to be sure we had not missed a trick. We then setup with our customer a plan whereby major and minor defects were properly defined. The gaging and testing were discussed and a common ground was found for

checking that assured our customer and ourselves that what was to be done inspection-wise would be the same in both places.

Think how valuable this is to companies doing business on this basis. When trouble arises you know exactly what the customer is referring to, what exact gages or tests are faulty and corrections can be made with a certainty that you are covering the right spot. When this method is compared to the many haphazard ways of the past the contrast is such that no one having had it would ever consider reverting to the old methods. If for no other reason, the intangible of mental strain reduction for both parties is more than worth the effort. This has been proven by our results generally plus a big reduction in returns of unsatisfactory goods.

Some time later we were able to certify, from another of our plants, 3 different air cleaners, one of the wet variety and two of the dry. This also enabled us to show savings of a very substantial nature.

Our fifth product that hit a Quality Level satisfactory enough to warrant certification was a brake lever. Again Quality Control played a major roll in making certification and savings possible.

At present we are working toward this same end on three other products. We expect to have these parts ready for certification within a very short time. One as a matter of fact is about ready now.

To sum up our experience we feel in analyzing all the features of Quality Control and Certification from both our angle and our customers that it has been most worthwhile.

Thank You

Well, Gentlemen, I have told you about the approach that we at Ford took when we approached our partners, as we like to call our suppliers, regarding our Quality Control Program and you have heard Mr. Spath and Mr. Rosengren tell how their companies felt when they were contacted. So I would like to close this meeting by saying that to us the Vendor's Response to the Quality Approach was exactly what we had hoped that it would be. Our suppliers were just as interested as we were in any new tool that will aid in improving Quality and reducing costs.

SOME BASIC PRINCIPLES OF SAMPLING

By

W. Edwards Deming
Bureau of the Budget
and
The Graduate School of Business Administration
New York University

The sampling of materials is a necessary part of the statistical control of quality. There can be no measurement of the quality of a lot of manufactured product unless we can somehow draw a proper sample for test. There is no point in developing elaborate methods of chemical analysis and of physical weighings and tests of samples of a material unless we know with assurance that the tests on the sample or samples will give us numerical values that are acceptable for the lot that the samples supposedly represent.

Before going on, let me differentiate between the theory of acceptance sampling and the problem that I have in mind, viz., sampling to estimate the average quality of a lot. In acceptance sampling we do not estimate the quality of a lot. That is not the problem. Instead, there is to be a succession of lots; a sample (let us confine ourselves to a single-sampling plan) is to be drawn from each lot, and the lot is to be accepted or rejected on the basis of the tests of the sample drawn from that lot. The rules for the size of the sample, and the rules agreed upon in advance for acceptance or rejection, constitute the sampling plan. The sampling plan is chosen so as to hold certain risks within prescribed limits at minimum cost, or to hold the cost to a certain level and to achieve the best protection possible at that cost. There is no talk of estimating the quality of a lot, nor in fact of estimating anything else, although a long series of samples will certainly give us as a by-product a deal of information that we may use in various ways, such as to estimate the process average.

The problem that I have chosen here is something different--something largely neglected in fact--; viz., (a) how to draw samples from a lot, in order that the theory of acceptance sampling may be applicable; and (b) how to compute the required size of sample in order to estimate the average quality of a lot, when such estimation is desirable.

This paper will not be an exhaustive treatment on sampling. Instead, it will purposely be limited to a few simple basic ideas. I choose to limit the scope of this treatment because in my experience the real needs of industry lie at present not in solutions of complex sampling designs, but rather (a) in getting started on some simple designs, and (b) in developing an appreciation for the fact that good sampling may pay high dividends.

By good sampling I do not necessarily mean high precision. I do mean precision controlled in advance to meet a need--to meet it and only to meet it. Extra precision is money wasted. On the other hand, insufficient precision may result in over- or under-payment on a single

lot, but with proper sampling plans (those that use random numbers, vide infra) such losses and gains are random, and in a long series of weighings or tests they will nearly cancel each other off, leaving only a small net balance one way or the other, fractional with respect to the savings accomplished by the introduction of sampling. In contrast, the losses or gains that result from the biases of judgment selections are systematic and irretrievable; they do not cancel, but pile up with time and may soon outweigh the cost of a statistical sampling plan.

Without the theory of probability it is impossible to control economically the precision of a sample to a desired level, or to eliminate bias. We must disregard convenience and follow rigidly the dictates of the theory; we shall have no right to use the theory to calculate the precision of a sample.

The basic requirement in statistical sampling is drawing only after thorough mixing. Otherwise the theory of probability has no chance to work for us. But, of course, most products can not be mixed; few will stand a jolt and jumble test. How then may we use statistical sampling? The most practical answer lies in the use of random numbers.

Randomness is not simply a haphazard arrangement. A random drawing is not simply one that is made without intentional bias. Randomness is an accomplishment, an achievement; it is the deliberate and intentional production of a required degree of disorder. We dare not assume that a manufactured product is thoroughly mixed. We must provide the randomization.

For an illustration of the use of random numbers, we may suppose that we have 970 articles in a lot and that we wish to draw 30 of them at random. How do we go about it?

Will it suffice to tell an inspector to take a sample by exercise of his judgment? Will it help to give him a few basic rules, such as to draw something from every part of the lot, not to choose the worst or the best, not to select the articles all off the front or all off the top? The answer is NO. Statisticians have positive proof that it is very dangerous to rely on any such instruction, no matter how carefully it be taught, and no matter how long or how successful a career the inspector has to his credit. Good intentions are necessary but not sufficient. I prefer to make better use of the inspector's knowledge. It is valuable: I don't like to throw it away.

The only way to draw a random sample is to invent some way to assign a serial number to each article. In the problem before us, the serial numbers will run from 1 to 970. These numbers need not be written on the articles. It is only necessary to prescribe a system of numbering--e.g., upward in the odd bays, downward in the even bays, starting at the front, and at the northwest corner. It is absolutely necessary that the system of numbering be unequivocal, so that if two inspectors seek article No. 167, they will both put their hands on the same one, without fail.

The next step is to read out from a table 30 random numbers between 1 and 970. Experience shows that men with no statistical experience can and will gladly learn to use a table of random numbers and to follow them faithfully, provided the men have some instruction on their use. There are now several good tables of random numbers. Turning haphazardly

to page 29 in Kendall & Smith's Random Numbers (Cambridge, 1951), I find these numbers, in order:

0794
8208
0991
7006
3115
Etc.

I agreed with myself in advance to cover up the column of figures at the extreme left. The remaining numbers then designate for the sample the articles that bear the serial numbers 794, 208, 6, 115, etc. I should continue onward until 30 articles have come into the sample. Why did I ignore the number 991? Because it is a blank: there is no such article. Any number that occurs twice, before I have the full sample of 30, is also a blank.

If we draw a sample in this manner, we shall have a sample that will be unbiased: losses and gains will approach 0 statistically in a long series of repeated weighings or tests, and the error of sampling, the precision of the average of any number of samples, can be calculated by the use of a simple formula.

In the matter of the assignment of serial numbers, it is sometimes more convenient to number the positions or the spaces that may contain an article or articles for inspection. This is especially true when the positions or spaces are already numbered in tiers of 10, 20, 100, etc.: serialization is then extremely simple. If a random number draws an empty space, it draws no item for the sample. If we had 1000 articles in 2000 spaces, we should only need to double the number of random drawings to yield a sample of about the required size. Or, curiously, if we have computed or otherwise decided that the ratio N/n (look ahead for definitions of these symbols) shall be 25, then we may still draw 1 space at random from each successive 25 spaces: the resulting sample will be very close to the required number of articles.

The random numbers may send the inspector to an inner layer. They may send him to a big article, to a small one, to low grade, to high grade. But so long as he follows the rules, and not his judgment, he will have a sample of known precision. That is, the sample will be one of a long series with a distribution whose mean is the value that we should obtain for a 100% sample, and whose standard deviation we may estimate well enough through use of the theory of probability from our one single sample of the series, if the sample is big enough.

A sample drawn with the use of random numbers will cost more than a haphazard collection of 30 articles drawn according to judgment. But it will be worth more. It will save more than its extra cost by providing us with extra protection, and by reducing the size of the sample required for that protection.

As a matter of fact, in looking at costs, one must also look at quality. There is no such thing as cost without a simultaneous measure of quality. This principle applies to anything, even to samples. The quality of the statistical sample is known; it is unbiased, and its precision can be calculated (vide infra). On the other hand, the usefulness of a judgment sample can only be inferred on the basis of judgment: hence

the name. Its quality is too often unknown, in which case no meaningful comparison of costs makes any sense.

Although this paper does not deal with acceptance sampling, sometimes I wonder how badly the producer's and consumer's risks are warped away from those in the tables by failure of the inspectors to draw samples at random. In my own experience I have seen severe biases arise from the failure to use random numbers. The fact that an inspector has no reason to select good articles or bad ones is no help: left to his judgment, he can only deliver biases. It is best to leave the selection to the random numbers.

The knowledge and the judgment of the inspectors can be put to use--in fact is indispensable--in the planning of a good sample, not in the selection of the particular articles for inspection. Later on I shall speak of the statistician's need of information concerning the handling of the material, its variability, the costs of handling it and of testing it, in order that he may draw up a satisfactory and workable plan. Then we shall see the need of all the help that we can get from the inspectors or from anyone else.

It is a simple matter to ensure that the sample is spread over the entire lot, without dependence on the random numbers to spread it. To do this we introduce the idea of stratification. It is only necessary to divide the lot into 30 zones or strata, approximately equal, and to draw one article from each zone. In this case, every 3d zone could be of size 33, the others of size 32. The inequality in probabilities is trivial, and our size of sample can be fixed in advance as 30. In the case of very unequal zones we should introduce blanks to create zones of equal size. A blank, if drawn, produces no sample.

There are many modifications of stratified sampling, but the main idea is contained in the above illustration. The Tukey plan (cited later) is one important modification. To employ it in stratified sampling one merely widens the zones by 10 and draws 10 articles from each zone.

Although stratified sampling has an appeal, theory shows that the gain in precision therefrom is often inconsequential. Sometimes people will waste money on stratification, without return, in sheer ignorance of the theory of sampling. There are formulas by which to express the expected gain in precision, and the statistician keeps them constantly in mind, along with costs, so as not to fail to use stratification when it will help, and so as not to use it if it only raises costs with but little compensating gain in precision. The fact is that the 30 random numbers, left to themselves, will give a pretty good spread over the lot, and that stratification pays a dividend only when we know that we shall encounter large differences between layers--large enough so that theory indicates substantial gains. A textbook on sampling should be consulted for further details. The intuition, if not trained by theory, is an expensive guide.

Although sometimes it is very expensive to introduce stratification into a sampling plan (in which case, as just explained, one should use theory to see if the extra cost is worth while), there are other circumstances in which it is actually as cheap and more convenient to do so, as when a lot of 1000 comes in multiples of 25, and we are to draw 1 in 25, or 1 in 50. Under such circumstances one should certainly use the strat-

ified plan. It may deliver a slightly better precision than an unstratified sample. With unimportant exceptions, it can not do worse.

To aim at a particular precision, one must use the theory of sampling, which is based on the theory of probability. The theory for the precision of a simple unstratified sample is not difficult to understand and to use. The standard error ($\sigma_{\bar{x}}$) of the mean of a sample is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation that we should find between tests if we were to test all the articles. For a more complex plan, one requires a more complex formula. If here we desire that $\sigma_{\bar{x}}$ shall be 5% of σ, we solve for n and find that n must be 400. If there are 10,000 articles in the lot, then for a stratified sample we should take 1 at random out of every 25.

If the sample-size n is 10% or 20% of N, the number of articles in the lot, and if we are fairly sure of σ, we may wish to take note of a slight reduction in the standard error, or of a slight reduction in the size of sample required. An easy way to remember this reduction is to compute N/n as $N(\sigma/\sigma_{\bar{x}})^2$, which in the present problem turned out to be 25, which indicates that we should take 1 article at random in 25. Strictly, however, the correct formula is

$$\frac{N}{n} = N\left(\frac{\sigma}{\sigma_{\bar{x}}}\right)^2 + 1$$

$$= 25 \qquad + 1 = 26$$

which gives 1 in 26, not 1 in 25. The reduction in this case is too trivial to bother about; and besides, 25 is an easier number to work with, especially if the articles come in arrays of some multiple of 25. On the other hand, if the first term on the right turned out to be 4, we should raise it by 1, and take 1 article in 5.

Although we speak of a sample of 1 in 5, or of 1 in 25, we really mean this strictly only if it is convenient to draw 1 from the first 5, 1 from the second 5, etc., or if theory indicates that, convenient or not, it will pay to do it that way. For an unstratified sample, a calculation of 1 in 5 is to be interpreted as 100 in 500; 1 in 25 as 100 in 2500, as the case may be, depending on the size of the lot.

It is important to note that the precision of a sample depends on the absolute size n of the sample, and hardly at all on its ratio to N. We may express the size of the sample in the form N/n, but this form is only for convenience. The size N affects appreciably the precision or the size of the sample only if n is 10% or more of N. The size N of the lot comes into the formula by the addition of unity on the right-hand side of the last equation.

Now we come to an evaluation of the precision of a valid statistical sample after it has been drawn and tested. What one is really interested in is not so much the precision aimed at (the intent of the plan), but the precision actually delivered (the quality produced). The most sig-

nificant characteristic of modern sampling is the fact that the sample itself furnishes an index of the precision actually delivered.

In this connexion, I may mention the Tukey plan for expediting calculation of the precision of a statistical sample after it is drawn and tested. This plan is especially useful if it is built into a complex plan, where calculation of the sampling error would otherwise become arduous beyond endurance. The Tukey plan is described in my book Some Theory of Sampling (John Wiley & Sons, 1950). It is really very simple: it consists of drawing 10 independent samples, each a 10th of the whole, by whatever plan we decide upon, however simple or however complex. Afterward, we calculate the sampling error in a jiffy by looking at the 10 different results. As a matter of fact, I now use the range of the 10 results, and not their mean square described in my book. To estimate the standard error of the mean it is only necessary to calculate R/10 where R is the range of the 10 results. The factor 10 is an approximation for $d_2 = \sqrt{10}$, which is more exactly 9.84. By drawing the sample on this plan, I was able to compute the 8 required standard errors on a recent job in about 3 minutes.

I may now turn to some of the complexities in sample-design. Some of them appear difficult, but melt away in consultation on the aims and requirements of the sampling plan. Others involve administrative decisions at a high level, so as to change a system of handling a material, or a system of records, in order to gain back later on the extra expense of these changes through more reliable and smaller samples.

There is the problem of new material, which frightens many people. In the calculations given a while ago on sample-size we used $\sigma_{\bar{x}}$ only in the ratio $\sigma:\sigma_{\bar{x}}$. Sometimes we wish to express $\sigma_{\bar{x}}$ in advance as so many thousandths of an inch. In such cases we need to know σ in the same units. Fortunately, it is always possible to discover a sufficiently firm value of σ to permit sensible decisions to be made on the size of sample required. In my own experience I meet this problem nearly every day. Questions directed to people who know the material that is about to be sampled will be able to give answers that the statistician may translate into σ. Here is where the experience and judgment of the inspectors can be put to use.

I have never been disappointed in the following approach. Ascertain roughly the lowest value that could exist in the lot; call it a. Ascertain roughly also the highest value; call it b. (I have never found this impossible.) Then assume that

$$\sigma^2 = \frac{1}{12}(b - a)^2$$

In this formula the distribution of measured values is assumed to be uniform from a to b. Perhaps the same sources of information that furnished rough values of a and b can tell us also that the distribution is roughly triangular, pointed upward near the middle, sloping off to near 0 at a and b. If so, take 1/24 in place of 1/12. If the distribution is roughly triangular with the base at either end, take 1/18 in place of 1/12.

These fractions represent about the extremes that one will meet in

practice. The factor 1/12 is the most conservative. That is, it is sure to deliver enough precision, or more. If the cost of testing the sample is a more serious matter than to fail to achieve a desired precision, it is well to take the factor 1/18. There is further discussion on this point on pages 62 and 108 of my book, mentioned by title earlier.

There is little excuse for being unaware in advance of the existence of some curious distribution that possesses a variance bigger than $\frac{1}{12}(b - a)^2$. If such a distribution appears to be likely, it probably arises from two distinct sources. The thing to do is to break up the material into two parts, for a stratified sample, and to sample the two strata separately.

In any case the precision of the sample once it is tested is not a matter of guesswork: it can be calculated, and in good statistical work this is always done. I shall speak more of this later.

Before we leave the subject of new material, let us be aware of the fact that every test of the new material builds up a background of experience for readjustment of the size of the next sample. Ignorance disappears in a hurry. By the time we have accumulated 20 tests carried out on a sample properly drawn, we have a great deal of information for future use. It is important to know that if our samples are not properly drawn, experience does not build up, and we remain in ignorance.

With many industrial materials it is exceedingly difficult to find ways to define any useful sampling unit; or to discover a way to cut costs so that any unit designated by the random numbers may be pulled out for test. It would be folly, for example, to talk of dividing up a pile of coal into cubic feet, and of drawing out cubic feet numbered 67, 121, 203 for test. But a pile of coal was not always in a pile. At one time or another it was on a conveyor-belt, or in wagons, in which state sampling units can be designated and drawn by a random procedure.* It would be folly also to designate for the sample a particular bale of burlap or wool, or a particular barrel of naphthalene, when these things are stored and piled up in a warehouse. We must face the facts. But again, these things were at one time in motion, and there may have been a good opportunity to draw a proper sample. It is being done.

A firm intent to discover a satisfactory procedure has never failed to bring forth results. Problems are not solved by taking refuge in the fact that the other man's problem was different from ours.

Under the circumstances that delineation of definite sampling units is possible, it is really very simple to design a foolproof but perfectly valid sample. I described one a while ago in a few lines. As a matter of fact, with a few simple modifications, such a plan is often the most economical one. On the other hand, such simple plans are often impracticable, being too expensive.

The really difficult problems in sampling are in the nature of com-

---

* There are many papers that illustrate this type of plan. I can not attempt to present a complete list. One of the latest and best is by Jack Hebden and Geoffrey H. Jowett, "The accuracy of sampling coal," Applied Statistics, vol. 1, Nov. 1952: pp. 179-191.

plex but foolproof designs for cutting costs, as by the painless introduction of a modified method for handling a material, or for keeping records, or by clever ways of using existing methods; by use of a 2-stage design (whereby we need delineate, number, and draw sampling units from only selected bays, layers, cases, or bales) when it will actually save money and not be too complicated; by use of a ratio in the formula of estimation; and other ways. A combination of perseverance, theory, ingenuity, and ability to use other people's knowledge, can sometimes solve hitherto insoluble problems. Anyone can devise an expensive foolproof sample that will meet any prescribed requirements of precision. But not everyone can design a sample that will be good enough, foolproof, and yet cheap enough. Here is where you tell the men from the boys. Neither experience alone nor theory alone will suffice.

Another test of statistical ability comes in patching up a sampling job that went wrong--some units not numbered, one number assigned to two units, an entire bay omitted and discovered later, and other deviations through misunderstanding of the instructions. There is always a way out, and without the introduction of bias. Sometimes the solution is simple, but one must know what to do.

Some additional discussion is contained in a paper by the author "On the sampling of physical materials," Revue internat'le de statistique, 1950: pp. 1-20; and in the references cited therein.

I shall close with the statement that Committee E-11 of the A.S.T.M., under the leadership of its chairman Harold F. Dodge, has a Task Group at work on the preparation of minimum standards for the sampling of materials, and for study and dissemination of the problems encountered in sampling. Several papers have appeared individually and collectively from this Task Group.

It may be of interest to know, too, that some of the most important scientific advances of the decade have been made in the sampling of business establishments, farms, households, people, and records, for current information of vital importance to business development and marketing. Standards for samples of this nature have been issued by the Office of Statistical Standards, with the aid of government statistical agencies, whose statisticians have contributed heavily to the advancement of statistical techniques in industry as well as in government. These standards of statistical surveys can be obtained on request to the Bureau of the Budget. The same theory and principles contained therein are applicable to the sampling of physical materials: the symbols don't care about the material.

BACKGROUND OF EXPERIMENTAL DESIGN

Martin A. Brumbaugh
Bristol Laboratories Inc.

Many millions of dollars are being expended today in experimental laboratories. A very brief study of the situation will show that constructive results are directly related to the amount of planning that precedes actual experimental work. One would run little risk of dispute in stating categorically that unless the planning stage is sound there is major risk of failure. However, much of the planning is done without the aid of a powerful modern tool -- the statistical approach to experimental work.

The statistical approach is not something different from the time honored scientific method, but merely the implementation of the scientific method by means of a set of objective procedures which guide the work toward the defined goal. It is a tool of investigation whose primary contribution is the integration and expediting of the work. It permits the investigator to see as a single unit the steps of preliminary planning, design, analysis, and interpretation. The resulting gain is not confined to the satisfaction of the investigator's scientific urge, but includes as well the desirable objective of more usable results per research dollar expended.

The ABC's of Statistical Design: There are ten distinct consecutive steps in statistical design.

I. Complete statement of the problem.

This requires time to secure complete agreement by all parties concerned as to what is to be investigated. The final statement may not have much resemblance to an initial assignment. It is extremely important that this statement be complete and that in that form it be approved by immediate superiors and by those who hold the purse strings.

II. Tabulation of available information on the problem.

The first step is to gain access to whatever is now in print concerning the problem. In some cases a usable solution will be discovered, thus eliminating the need for additional investigation. In any case, all pertinent information should be recorded and kept available. The second step is to obtain from colleagues and from acquaintances outside the organization any direct or collateral information on the problem.

III. Digest of information and preparation of preliminary plan.

With the problem completely defined and all available information in hand the investigator is ready to design an experiment. In some cases the design will be simple and obvious, but in most cases the knowledge of the statistician is required, because he is trained to foresee the consequences analytically of a plan and to

introduce maximum economy of experimental work without sacrificing any essentials. The crux of his contribution is his knowledge of the principles of variability and his understanding of how to operate the experiment so that the effects of the important factors can be segregated for individual evaluation.

The initial design may not finally be put into operation. It should be the subject of very searching inquiry by the investigator and the statistician: this statement continues to be true whether two individuals are involved or one. In some cases a pilot run of all or of parts of the plan may be possible. If so, shortcomings may be discovered, correction of which will improve the plan ultimately put into operation.

IV. Enumeration of all possible results from the design.

This is peculiarly the field of the statistician. He should follow through the analysis to consideration of questions such as the following: If the average results for factor A are significantly different, what does this mean in terms of the experiment? If an interaction between factors A and B appears what does this mean? If the error term is very large or very small what does this mean? What is measured by each source of variability? The questions to be considered depend upon the design of the experiment.

Another question should be considered in advance of actual experimental work. Many times some unforeseeable circumstance will interfere with the experimental work. For example, a factor in the experiment may be analyst difference, but midway through the work an analyst becomes ill and must withdraw. Provision should be made in advance for this contingency. There are others of a similar nature. The objective should be to plan so that improvisation in the course of the experimental work will be eliminated.

V. Record the interpretation to be given to each anticipated result and the argument required to sustain such interpretation.

The purpose of this step is fairly obvious. In the course of the investigation results will be obtained which may appear to support one hypothesis or another. It is difficult to retain an unbiased attitude toward these unfolding results. When the analysis has been completed the preliminary bias tends to creep into the final interpretation. The most powerful device for avoiding this bias is to have available a complete interpretation prepared before the experimental work was undertaken. A further obvious advantage comes from the existence of a written record of earlier thinking, if a disagreement on interpretation should arise between the investigator and the statistician. One can scarcely expect that by this device all disagreement will be avoided

but it certainly will eliminate the "you did; I did not" phase of the discussion.

VI. Revision of the design, if indicated by IV and V.

The entire plan can now be evaluated. If the interpretation of possible results indicates weakness or perhaps inefficiency in the design, such shortcomings should be remedied. Any major change at this point will delay the experimental work, but the improvement may produce an overall gain or lower cost or both. One of the major requirements at this point is flexibility of thinking by investigator and statistician. Too often an inefficient experimental program is put into operation because those responsible take the position that it is too late to change the plan.

VII. Arrangements for experimental work.

At this point the emphasis shifts to a somewhat different type of planning. Proper equipment must be requisitioned from storerooms or perhaps purchased. The experiment must be fitted into the laboratory schedule. Participants must be instructed in detail as to their work. Forms for record keeping must be prepared so that they become part of the instruction. A time schedule must be set up. If conditions permit, a preliminary run of some part or possibly all of the experiment should be made. Analysts can fail to follow instructions no matter how carefully the instructions were presented.

An extremely important part of these arrangements is proper provision for taking samples. The design of an experiment has often been destroyed through failure of operators of a chemical process to take samples as directed. This failure can be expensive in any experiment, although a chemical experiment probably is most vulnerable on this score.

VIII. Conduct of the experimental work.

The statistician does not participate in this step unless an emergency arises. Despite the best planning, things can go wrong. Joint action of investigator and statistician are required to deal with this contingency. The investigator must consider the effect of the situation on laboratory procedures, the independence of results, bias, etc. The statistician must consider the effect of the situation on his ability to carry out the planned statistical analysis of results.

Such things as damage to equipment, contaminated reagents, loss or breakage of a sample, breakage of tools, sickness of animals and similar things may force reoperation of one or more blocks of an experiment. In doing this, great care must be exercised to avoid introducing a new source of variability into the experiment. An example of this

sort came to the author's attention recently. A clinical experiment with a new drug was carried out in a hospital far removed from the base of the investigator and statistician. When the results were analyzed, the status of the drug would have been quite clear, if it had been possible to verify the suspicion that with one group of patients the control drug and the new drug were switched. No verification of this suspicion was possible, hence, the experiment was repeated. The results of the second experiment were consistent in every detail with the first, provided a switch in results could be made for the one group of patients. It should be added that this was not a large experiment, yet several thousand dollars were involved in repeating the work.

IX. Analysis of results.

This is the field of the statistician. If his preliminary work has been well done, the analysis becomes largely a routine operation. Certain details such as preparation of forms, provision for checking of calculations, and labeling to avoid confusion, should not be overlooked. An omission which can be a source of trouble is failure to edit the data before starting the analysis. The statistician should personally study the data to detect inconsistencies, omissions, irregularities, etc. Anything discovered should be clarified prior to analysis.

X. Interpretation of results and presentation of a report.

The pattern of the interpretation has been established at step V. The results of analysis should fall into the pattern. Perhaps additional information will appear; suspected relationships will either be confirmed or rejected. The great advantage of the preliminary interpretation resides in avoiding bias. It is all too easy to misinterpret a result when all of the data are at hand and all of the relationships of the factors have been studied intensely during the analysis. Reference to the reasoning set down before the experimental work was undertaken is a strong deterrent to biased thinking.

The preparation of a report is of first line importance. This report must cover the method of investigation, the experimental procedure, and the results obtained. Whether the report is a memorandum to a supervisor, a general communication within the organization, or a published document, the rules are the same; tell the story completely, efficiently, and attractively. This advice is usually distasteful to the investigator, who is wholly engrossed in scientific pursuits and certainly has a minor interest in literary accomplishments. Yet the history of science is replete with cases of investigative work of a high order which went unrecognized because of faulty presentation.

ON THE DESIGN OF EXPERIMENTS

D. B. DeLury
Ontario Research Foundation

This discussion is intended to provide a glimpse of the thinking that goes into the planning of experiments and to explain the use of some simple tools that are helpful in translating these ideas into effective experimental designs. The discussion is conducted in terms of a single example, but the approach is entirely general.

Let us suppose that we wish to compare two types of ammunition which are supposed to perform alike but which differ enough in methods of manufacture or in other ways that it is meaningful to ask if they are, in fact, essentially the same. Let us say that we are concerned only with the accuracy of the ammunition and that we have an acceptable method of measuring accuracy.

Proposal I

We might take a number n of cartridges of each of the two types (type 1 and type 2), fire them all from the same rifle and measure the accuracy of each cartridge. The cartridges of type 1 furnish n numbers, $x_{11}, x_{12}, \ldots x_{1n}$ and those of type 2 yield $x_{21}, x_{22}, \ldots x_{2n}$. We wish to decide, on the basis of these two sets of numbers, if there is evidence of a real difference between the two types of ammunition.

The decision would be easy if all the x_1's were equal and all the x_2's were equal. Presumably this would never happen and our decision must be reached in the face of variation within each of our two types.

We should perhaps pause for a moment to realize that observations are always interpreted in terms of a model. What sort of model would we construct to reflect a possible difference between the two types of ammunition and also variation within each type?

In order to cope with variation within types, we envisage the accuracies of all the ammunition of one type assembled in the form of a frequency distribution. This distribution is characterized by a mean, a variance and perhaps other features, but if the distribution is normal the mean and the variance specify it completely. We might express the value of any member x of this distribution as

$$x = \bar{x}' + e,$$

when $\bar{x}'$ is the mean and e (the "error") is a member of a distribution identical with that of the x's except that its mean is zero.

In terms of such a model, then, each observation may be expressed in the form

(1) $x_{11} = \bar{x}_1' + e_{11}$, i = 1, 2, ... n, (type 1);

$x_{21} = \bar{x}_2' + e_{21}$, i = 1, 2, ... n, (type 2).

Our question is: can we decide whether or not $\bar{x}_1' = \bar{x}_2'$, even though we do not know the values of any of the e's?

Several considerations enter into the answering of this question. In the first place, we cannot say anything about the e's except in terms of the distribution from which they came. It is therefore of crucial importance that they come from the distribution in such a manner that the pattern of the distribution be not persistently misrepresented. This is the reason for the insistence on random selection of samples.

Secondly, unless we know more about these distributions than we usually do, we must make provision, in the taking of the observations, for learning something of the nature of the distributions we are dealing with. The sampling suggested under Proposal I evidently does furnish information of this kind. Indeed, if n were large enough, we could arrange each of the two sets of observations in a frequency distribution which would, if the samples were randomly chosen, reflect the properties of the distributions from which they were taken. Sometimes it is necessary to take large enough samples to do this. Often, however, we have grounds for assuming that our distributions are not seriously different from the normal type. If this can be granted, it is sufficient simply to estimate the mean and the variance. Smaller samples are adequate for these purposes.

The method that will be adopted here for extracting this information from the samples is one that generalizes easily to more complicated experiments. Let us say, for our illustration, that n = 4, so that 4 cartridges of type 1 and 4 of type 2 were tested. The method of analysis consists in transforming the 8 numbers yielded by the test into 8 other numbers by a transformation chosen to put in evidence the quantities that we wish to examine and also to possess certain other desirable properties that will be discussed later.

The transformation that will be used in this example is shown in Table I, written in a form that avoids as much writing as possible and that displays clearly the important features of the transformation.

Table 1

	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	Divisors
y_1	1	1	1	1	1	1	1	1	$\sqrt{8}$
y_2	1	1	1	1	-1	-1	-1	-1	$\sqrt{8}$
y_3	1	1	-1	-1	1	1	-1	-1	$\sqrt{8}$
y_4	1	1	-1	-1	-1	-1	1	1	$\sqrt{8}$
y_5	1	-1	1	-1	1	-1	1	-1	$\sqrt{8}$
y_6	1	-1	1	-1	-1	1	-1	1	$\sqrt{8}$
y_7	1	-1	-1	1	1	-1	-1	1	$\sqrt{8}$
y_8	1	-1	-1	1	-1	1	1	-1	$\sqrt{8}$

This array is to be interpreted to mean that each of the y's is to be formed by multiplying each coefficient in its row by the x at the top of the column which contains the coefficient, adding these products and dividing by the divisor listed at the end of its row. For example,

$$y_4 = (x_{11} + x_{12} - x_{13} - x_{14} - x_{21} - x_{22} + x_{23} + x_{24})/\sqrt{8}$$

Several facts about this transformation may be remarked at once.

1. The 8 x's are converted into 8 y's, so that nothing contained in the values of the x's is lost when the x's are replaced by the y's.

2. When the expressions for the x's that we assumed in our model (equation I) are substituted in the transformation we find:

$$y_1 = \frac{4}{\sqrt{8}}(\bar{x}'_1 + \bar{x}'_2) + \text{combinations of the e's;}$$

$$y_2 = \frac{4}{\sqrt{8}}(\bar{x}'_1 - \bar{x}'_2) + \text{combinations of the e's;}$$

$$y_3, y_4, \ldots y_8 = \text{combinations of the e's.}$$

We may agree that y_1 exhibits values in which we are not interested. y_2, however, contains the difference between the means, whose evaluation is the chief object of the test. The rest of the y's express contributions of error only. Thus it is seen that the transformation has accomplished, in some measure, a separation of the two sorts of variation that we envisaged in our model.

If we allow another assumption at this stage, we can go farther.

Assumption

The two distributions are normal and have equal variances.

The assumption that the variances are equal could be subjected to a test and so also could the question of normality if the samples were large enough. We shall not delay over these tests here, but shall proceed on the supposition that the conditions of normality and equality of variances are not seriously violated. It is then a simple exercise in probability theory to prove 3. The proof is not given here.

3. The y's are distributed in independent, normal distributions, with the same variance as that of the x distributions.

One more result of probability theory will be stated without proof.

Let $v_1, v_2, \ldots v_p, w_1, w_2, \ldots w_q$, (p and q any positive integers) be independent, normally-distributed variables with zero means and equal variances. Then,

$$F = \frac{(v_1^2 + v_2^2 + \ldots + v_p^2)/p}{(w_1^2 + w_2^2 + \ldots + w_q^2)/q}$$

has a known distribution (called the analysis of variance distribution or the F distribution). The probability that F should lie in any given range can therefore be calculated and tables are available to assist in this calculation.

Let us return now to the example. We know that, under our assumptions, the values of $y_3, y_4, \ldots y_8$ may be regarded as observations on independent, normally-distributed variables with zero means and equal variances and that, if $\bar{x}'_1 = \bar{x}'_2$, y_2 satisfies the same conditions. Therefore, if $\bar{x}'_1 = \bar{x}'_2$, the ratio

$$\frac{y_2^2/1}{(y_3^2 + y_4^2 + .. + y_8^2)/6}$$

may be viewed as an observation drawn from the F distribution. On the other hand, if $\bar{x}'_1$ differs from $\bar{x}'_2$, this ratio will run to bigger values than would be predicted from the F distribution. This furnishes the basis for a test of the hypothesis $\bar{x}'_1 = \bar{x}'_2$. If the value of the ratio, calculated from our samples, attains a value so large that it would be reached only rarely by an F ratio, we would attribute its large value to a non-zero value of $\bar{x}'_1 - \bar{x}'_2$.

The transformation that converted the x's into the y's is evidently of crucial importance in this development. The fact, stated above, that the set of independent, normally-distributed x's with equal variances transformed into a set of independent, normally-distributed y's having the same variance as the x's, depends on a property of the transformation called orthogonality. This property is ensured by choosing the coefficients of the transformation to satisfy two conditions.

Suppose that one of the y's, y_i say, is given by

$$y_i = a_1x_{11} + a_2x_{12} + \ldots + a_8x_{24}$$

and that another y, y_j say, is given by

$$y_j = b_1x_{11} + b_2x_{12} + \ldots + b_8x_{24}.$$

Then, to meet the requirements of orthogonality,

(a) $\quad a_1b_1 + a_2b_2 + \ldots + a_8b_8 = 0.$

This condition must hold for every pair of y's in the transformation.

(b) $\quad a_1^2 + a_2^2 + \ldots + a_8^2 = 1.$

This condition must be met by every y in the transformation.

Evidently it is always possible to write an orthogonal transformation of any number of x's into the same number of y's. Indeed, there

are infinitely many such transformations. Our concern is whether one can be found that will segregate the systematic variations that are present in the data into independent components. This is not always possible and it is one of the aims in planning the experiment to arrange that it shall be possible.

In an experiment conducted according to Proposal I, the one possible systematic variation envisaged in the model is the difference between the two means, $\bar{x}'_1 - \bar{x}'_2$. We have seen that any contribution from this source is neatly parcelled into y_2 and appears nowhere else. Therefore the remaining y's (except y_1) are left free to display random variation only. The part of the transformation specifying y_3, y_4, ... y_8 could be given in many different ways, but they would all be equivalent, for the uses we made of these components. To see this, one more simple consequence of orthogonality is needed: the sum of the squares of the y's is equal to the sum of squares of the x's.

$$y_1^2 + y_2^2 + \ldots + y_8^2 = x_{11}^2 + x_{12}^2 + \ldots + x_{24}^2.$$

This is easily verified by direct calculation.

Referring now to the transformation of Table I, we see that

$$y_1^2 = \frac{(Sx)^2}{8} = 8\bar{x}^2 \text{ , (Sx means the sum of the x's)}$$

$$y_2^2 = \frac{(Sx_1 - Sx_2)^2}{8} = 2(\bar{x}_1 - \bar{x}_2)^2.$$

Therefore,

$$y_3^2 + y_4^2 + \ldots + y_8^2 = Sx^2 - 8\bar{x}^2 - 2(\bar{x}_1 - \bar{x}_2)^2 = S(x_1 - \bar{x}_1)^2 + S(x_2 - \bar{x}_2)^2.$$

It is clear, from this relation, that the sum $y_3^2 + y_4^2 + \ldots + y_8^2$ is determined completely by the values of y_1 and y_2, provided the orthogonality conditions are met and that the values of y_3, y_4, ... y_8 need not, for our present purposes, be specified. Indeed, we see that it is not necessary to refer explicitly to the y's at all to compute the test-ratio, because it is simply

$$\frac{2(\bar{x}_1 - \bar{x}_2)^2}{\left[S(x_1 - \bar{x}_1)^2 + S(x_2 - \bar{x}_2)^2\right]/6}$$

which is recognized as the square of the t ratio usually used to test the difference between two sample averages.

It is customary and convenient to summarize the results of the calculations in an "analysis of variance table" like Table 2. In ordinary use, of course, numbers are entered instead of formulae. The

entries in the "degrees of freedom" column are simply the numbers of y's squared and added together to form the "sum of squares".

Table 2

Source of Variation	Degrees of Freedom	Sum of Squares
Between samples	1	y_2^2
Within samples (error)	6	$y_3^2 + y_4^2 + y_5^2 + y_6^2 + y_7^2 + y_8^2$

The test-ratio is easily calculated from the numbers in the analysis of variance table and it may, if desired, be entered in the table. If the test-ratio is significantly large, we can exhibit the difference $\bar{x}_1 - \bar{x}_2$ with some confidence that it reflects something more than error; if the ratio is not significantly large, there is a reasonable chance that error alone could cause so large a difference. This does not prove, of course, that there is no difference. We may have failed to detect a real difference because our error was large. With a view to cutting down the magnitude of the error, we might argue along the lines of Proposal 2.

Proposal 2

Let us suppose that the powder that goes into these cartridges is produced in batches and that it is not unreasonable to fear that the accuracies of the cartridges are affected by batch-to-batch differences in the powder. This possibility was ignored in the sampling conducted under Proposal I and batch-to-batch differences may have added their systematic contributions to the terms that we attributed to error. This could be avoided, of course, by making sure that all the test cartridges were made with powder from the same batch, but this arrangement would yield a comparison of types of restricted value, since there would be no assurance that similar results would be found with other batches. It is obviously preferable to spread the comparison of types over several batches in a controlled fashion, so that differences between batches are not allowed to masquerade as error.

For our example, let us bring only two batches into the test and select randomly two cartridges of each type from each batch. The test results could be presented in a two-way table.

	Batch 1		Batch 2	
Type 1	x_{11}	x_{12}	x_{13}	x_{14}
Type 2	x_{21}	x_{22}	x_{23}	x_{24}

The model we must use to interpret these observations is necessarily more elaborate than that used under Proposal I, because we must provide for a possible difference between batches and also we must be prepared for the possibility that the difference between the two types

of cartridge may not be the same in batch 2 as it is in batch 1.

If we employ the transformation of Table 1 on these observations, inspection shows that

y_2 expresses the overall difference between types, plus error;

y_3 expresses the overall difference between batches, plus error;

y_4 expresses the difference, between one batch and the other, of the difference between types;

y_5, y_6, y_7, y_8 express error only, as reflected by the differences between pairs of the same type from the same batch.

The analysis of variance table now takes the form of Table 3.

Table 3

Source	d.f.	s.s.
types	1	y_2^2
batches	1	y_3^2
types x batches interaction	1	y_4^2
within pairs (error)	4	$y_5^2 + y_6^2 + y_7^2 + y_8^2$

Simple rules for calculating the sums of squares directly from the observations can be developed from the transformation. The details will not be given here.

Test ratios can be set up as before. Naturally one would first set up the ratio

$$\frac{y_4^2}{(y_5^2 + y_6^2 + y_7^2 + y_8^2)/4}$$

to test the interaction of types and batches, because if the difference between types varies from batch to batch, no clear-cut meaning can be attached to "difference between types" as displayed by y_2 and there is no occasion to test y_2. On the other hand, if y_4 is not significantly large, one would conclude that the difference between types is reasonably constant from one batch to another and it would be meaningful to test y_2.

Proposal 3

The comparisons made under Proposals 1 and 2 were carried out using one rifle, but it is not likely that we would be satisfied to assume that the difference between types would be the same in all rifles and we would prefer to spread our test over a number of rifles. Let us say, to keep our example small, that 2 rifles are to be brought into the

test. One of each pair of cartridges of the same type and made from the same batch of powder could be assigned to one rifle, the other of the pair to the other rifle. The results of such a test would naturally be entered in a three-way table.

	Batch 1		Batch 2	
	Rifle 1	Rifle 2	Rifle 1	Rifle 2
Type 1	x_{11}	x_{12}	x_{13}	x_{14}
Type 2	x_{21}	x_{22}	x_{23}	x_{24}

When the transformation of Table 1 is applied to these observations, we see that, in addition to error,

y_2 expresses the overall difference between types;

y_3 expresses the overall difference between batches;

y_4 expresses the types x batches interaction;

y_5 expresses the overall difference between rifles;

y_6 expresses the types x rifles interaction;

y_7 expresses the batches x rifles interaction;

y_8 expresses the types x batches x rifles interaction.

None of the y's expresses only the contribution of error; this arrangement makes no provision for estimating error. In some circumstances this may not be a serious defect, but usually it is. In any event, let us take the position that an estimate of error is essential and enlarge the experiment to provide for it. We might think of selecting another set of 8 cartridges like the first set of 8, to provide duplicates which would yield an estimate of error in the same manner as they did under Proposal 2. However, let us, for illustrative purposes, take the view that it is not possible to test more than 8 cartridges at one time under uniform conditions. We shall therefore run two separate tests, each under uniform conditions but with the possibility that conditions may change somewhat between one test and the other. Estimation of error through differences between duplicates is now impossible. However, the fact that the level of accuracy may be different in the two tests does not stand in the way of assessing error. We are, in these tests, concerned only with comparisons of accuracies, that is, with differences and the error that affects these comparisons is displayed in the variation of the differences from test to test. This is precisely the variation that is gathered together in the "interaction with tests" components, which will therefore be defined as the "error" components. The two tests will be called "replications", since each of them includes a complete set of comparisons.

A transformation of the 16 numbers that would result from such an experiment is shown in Table 4. The symbol (i j k l) stands for the accuracy obtained with type i, batch j, with rifle k in replication l.

The headings Test 1 .... Test 4 and the dotted vertical lines refer to Proposal 4 and should be ignored at this stage. The solid line separates the two replications. The component y_1 has not been entered in the table and the divisors have been omitted, for lack of space on the page.

Table 4

		Test 1				Test 2			
		(1111)	(1221)	(2121)	(2211)	(1121)	(1211)	(2111)	(2221)
(types)	y_2	1	1	-1	-1	1	1	-1	-1
(batches)	y_3	1	-1	1	-1	1	-1	1	-1
(t x b)	y_4	1	-1	-1	1	1	-1	-1	1
(rifles)	y_5	1	-1	-1	1	-1	1	1	-1
(t x r)	y_6	1	-1	1	-1	-1	1	-1	1
(b x r)	y_7	1	1	-1	-1	-1	-1	1	1
(t x b x r)	y_8	1	1	1	1	-1	-1	-1	-1
(Reps)	y_9	1	1	1	1	1	1	1	1
(R x t)	y_{10}	1	1	-1	-1	1	1	-1	-1
(R x b)	y_{11}	1	-1	1	-1	1	-1	1	-1
(R x t x b)	y_{12}	1	-1	-1	1	1	-1	-1	1
(R x r)	y_{13}	1	-1	-1	1	-1	1	1	-1
(R x t x r)	y_{14}	1	-1	1	-1	-1	1	-1	1
(R x b x r)	y_{15}	1	1	-1	-1	-1	-1	1	1
(R x t x b x r)	y_{16}	1	1	1	1	-1	-1	-1	-1

Table 4 (Continued)

		Test 3				Test 4			
		(1112)	(1222)	(2122)	(2212)	(1122)	(1212)	(2112)	(2222)
(types)	y_2	1	1	-1	-1	1	1	-1	-1
(batches)	y_3	1	-1	1	-1	1	-1	1	-1
(t x b)	y_4	1	-1	-1	1	1	-1	-1	1
(rifles)	y_5	1	-1	-1	1	-1	1	1	-1
(t x r)	y_6	1	-1	1	-1	-1	1	-1	1
(b x r)	y_7	1	1	-1	-1	-1	-1	1	1
(t x b x r)	y_8	1	1	1	1	-1	-1	-1	-1
(Reps)	y_9	-1	-1	-1	-1	-1	-1	-1	-1
(R x t)	y_{10}	-1	-1	1	1	-1	-1	1	1
(R x b)	y_{11}	-1	1	-1	1	-1	1	-1	1
(R x t x b)	y_{12}	-1	1	1	-1	-1	1	1	-1
(R x r)	y_{13}	-1	1	1	-1	1	-1	-1	1
(R x t x r)	y_{14}	-1	1	-1	1	1	-1	1	-1
(R x b x r)	y_{15}	-1	-1	1	1	1	1	-1	-1
(R x t x b x r)	y_{16}	-1	-1	-1	-1	1	1	1	1

The analysis of variance table is given in Table 5.

Table 5

Source	d.f.	s.s.
types	1	y_2^2
batches	1	y_3^2
types x batches	1	y_4^2
rifles	1	y_5^2
types x rifles	1	y_6^2
batches x rifles	1	y_7^2
types x batches x rifles	1	y_8^2
replications	1	y_9^2
interactions with reps. (error)	1	$y_{10}^2 + y_{11}^2 + y_{12}^2 + y_{13}^2 + y_{14}^2 + y_{15}^2 + y_{16}^2$

The pattern of this experiment is commonly called the randomized block. "Block" here means simply a set of uniform conditions within which all the comparisons of our experiment can be tested. "Randomized" implies that all the observations within a block should receive error contributions drawn randomly from the same error distribution. We have already insisted on randomness in the selection of cartridges for the test. Perhaps we should also insist that, within each block, the cartridge-rifle combinations should be tested in random order. We should certainly insist on this if we feared that there might be some small systematic drift in conditions within the block, but even if none exists, it costs nothing to randomize and nothing can be lost by doing so. The practice should therefore always be followed. This is not to say that we should try to "randomize out" large systematic sources of variation that are known to exist. All that is implied is that randomization should take over where control leaves off.

It may be helpful at this stage to draw attention to some features of the transformations we have been using.

1. Each "main effect" component is formed by summing over all factors but the one whose levels are to be contrasted. It follows from the balance that exists in the arrangement of the experiment that the coefficients of this component sum to zero within each level of each of the factors which do not contribute to this main effect. For example, the coefficients in y_2, the "between types" component, sum to zero within each batch and within each rifle. This property is most important. If the coefficients of y_2 did not sum to zero within each batch, for example, y_2 would include batch differences as well as

differences between types. This brings out also the importance of the balance that has been arranged in the experiment. If the balance were disturbed in any way it would not be possible to find an orthogonal transformation that would separate the various kinds of variation into separate, independent components.

The summing to zero of sets of coefficients provides a useful test for finding out which sources of variation contribute to a given component and which ones do not.

2. The coefficients of the component that specifies the interaction of two main effect components can be obtained by multiplying corresponding coefficients of the two main effects. This rule simplifies greatly the writing of the transformation.

Proposal 4

Let us suppose now that it is found to be impossible to test as many as 8 cartridges under uniform conditions but that uniformity is possible with 4 cartridges. It is desirable, therefore, to split each replication into two tests, each of 4 cartridges. There are many ways of making the division of 8 into 2 sets of 4 and it is important to recognize the consequences of any allocation that may be proposed. For example, suppose that, in each replication, the cartridge-rifle combinations symbolized by (111), (122), (212), (221) are put into one test and the remaining combinations, (112), (121), (211), (222) into the second test. This division is shown by the dotted lines in Table 4. Inspection of the components in the table shows that all but two, y_8(t x b x r) and y_{16} (R x t x b x r) have coefficients that sum to zero within each test. It follows that any differences there may be between tests appear only in these two components, which are thereby sacrificed. If, therefore, we have reasonable grounds for believing that the types x batches x rifles interaction is unimportant, this arrangement provides an acceptable method of taking advantage of such uniformity as can be attained in the conditions of the test.

It is said, of this arrangement, that the types x batches x rifles interaction is confounded with tests. Evidently the sets could be selected so as to confound any one of the components with tests. Indeed, we might confound a different component in each replication.

The analysis of variance table for the confounded arrangement shown in Table 4 is given in Table 6.

Proposal 5

Let us suppose now that circumstances impose one more condition, that we cannot test (or prefer not to test) more than 8 cartridges altogether. Which ones, of the 16 used in Proposals 3 and 4, should be taken?

We could, of course, run one complete replication, as was mentioned under Proposal 3, but in that event there would be no provision for estimating error. Even so, experiments involving large numbers of factors are sometimes run in one replication, with the expectation of using high order interaction components for estimating error. This

Table 6

Source	d.f.	s.s.
types	1	y_2^2
batches	1	y_3^2
types x batches	1	y_4^2
rifles	1	y_5^2
types x rifles	1	y_6^2
batches x rifles	1	y_7^2
tests	3	$y_8^2 + y_9^2 + y_{16}^2$
error	6	$y_{10}^2 + y_{11}^2 + y_{12}^2 + y_{13}^2 + y_{14}^2 + y_{15}^2$

practice rests on the principle that the greater is the number of factors in an interaction component, the stronger are the general grounds for believing that it contains no systematic contribution from the factors and therefore reflects only error. This is obviously a weaker principle for estimating error than those used earlier in this discussion, which depend on explicit repetition. One would hesitate to employ it when as few as three factors are involved.

An alternative to running one complete replication is to run twice one of the two sets into which we divided a replication under Proposal 4. We might, for example, run Tests 1 and 3 of Table 4. If we think of obliterating the portion of Table 4 that contains Test 2 and Test 4, it will show what happens to the components when only Tests 1 and 3 are carried out. y_8 becomes identical with y_1, y_7 with y_2, y_6 with y_3, y_4 with y_5, y_{16} with y_9, y_{15} with y_{10}, y_{14} with y_{11}, y_{13} with y_{12}. Out of all this, we can find 7 different meaningful components, which are listed in the following analysis of variance table. It is to be remembered that the y's in Table 7 refer to Table 4 with Tests 2 and 4 removed and the divisors changed accordingly.

Table 7

Source	d.f.	s.s.
types + batches x rifles	1	y_2^2
batches + types x rifles	1	y_3^2
rifles + types x batches	1	y_4^2
replications	1	y_9^2
error	3	$y_{10}^2 + y_{11}^2 + y_{12}^2$

We see that the consequence of this arrangement is to entangle completely each main effect with an interaction. It is therefore an intolerable arrangement unless we have strong assurance that certain factors do not interact. When the number of factors is large, we can arrange that main effects are confused only with interaction components of high order that, in all likelihood, contain no systematic effects. Note that this is the same argument as was advanced earlier to support the use of high order interaction components to estimate error. Indeed, in many situations, there are few grounds for choosing between using high order interaction components for the estimation of error and this arrangement, which is called "fractional replication" or "fractional factorial". In the example we have constructed, neither plan would be acceptable.

QUALITY CONTROL PROJECT FOR 105MM AMMUNITION

Arthur Stein
Ordnance Ammunition Center, Army Ordnance Corps

I. Introduction

During the course of this talk I may refer to some words, descriptive of ammunition or its performance, which may be unfamiliar to some of you. These perhaps had best be defined at the outset. The item with which we are concerned is the 105mm high explosive complete round. Such a complete round is illustrated in Figure 1. As you see, it is an assembly consisting of a projectile, cartridge case, propellant powder and primer. The projectile is a shell which is loaded with high explosive and to which is attached a fuze. When the firing pin in the gun strikes the primer head, the primer flashes and ignites the propellant. The pressure created by the gases thus generated propels the shell through the gun tube. The soft metal rotating band is engraved by the rifling of the tube, thus rotating the shell to give it stability during its flight. The fuze is the brain of the shell and sets it off when it hits something. The speed with which the shell leaves the gun is called the muzzle velocity. The study of the actions taking place within the gun tube and chamber is called interior ballistics.

The study of the flight of the projectile through the air is called exterior ballistics. The distance the shell travels is the range.* When measured in a horizontal plane through the gun muzzle in the direction of the muzzle it is the horizontal range. Not all shell point in the direction they are moving. The angle between the longitudinal shell axis and the direction of flight is the yaw angle.

The 105mm howitzer complete round is semi-fixed, which means that the cartridge case is not crimped to the shell but instead may be easily pulled away from the shell to permit removal of one or more of the seven powder increments which make up the propellant charge. This enables greater flexibility in firing over obstacles, since one could obtain a desired short range at high elevation by removing a sufficient number of the powder increments.

Variations in shell weight which are permitted by the present specifications result in significant variations in range. In fact, this effect is so significant that it has been necessary to subdivide this total permitted weight variation into three weight zones, called Zones I, II and III. Shell in each of these three weight zones are kept segregated in separate lots up to the gun in the Field. The artilleryman essentially uses separate firing table values for each zone in determining his ranges.

To supply the best possible weapons to the field we need ammunition that is safe to handle, that functions as intended and possesses uniform velocity and flight characteristics. Safety and efficiency require a ballistically uniform product. Shell have to land where you think they

*Throughout this talk the words "maximum dispersion" will be used to designate the difference between the highest and lowest measurements in a series. The word "range" will be reserved for its special ballistic meaning.

105 MM H.E. COMPLETE ROUND

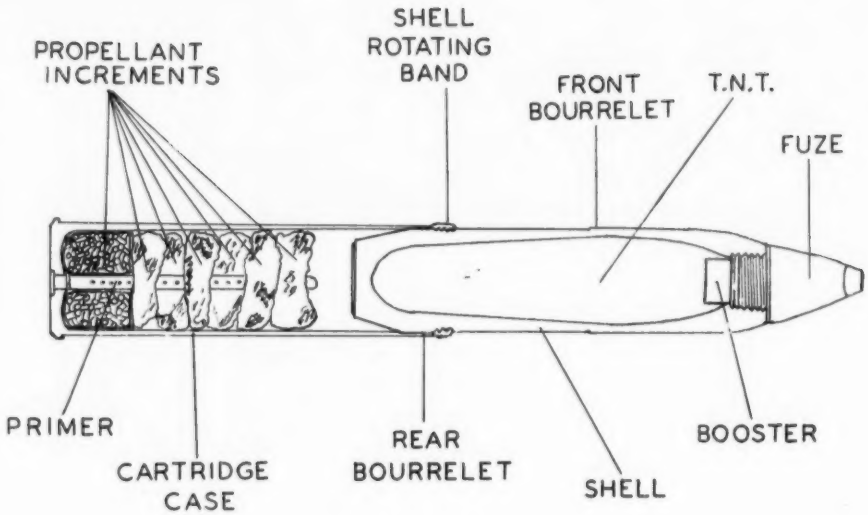


FIG. 1

are going to land, else they may fall on our own troops or require an unduly large expenditure to accomplish a given objective.

At the suggestion of Major General L. E. Simon, Chief of the Research and Development Division in the Office of the Chief of Ordnance, representatives of the Ordnance Corps met with Mr. G. D. Edwards of the Bell Telephone Laboratories and the first president of our society, to discuss a somewhat different approach to this problem of ballistic uniformity--to determine the characteristics affecting ballistic uniformity and to try to bring those characteristics into control. As a result of these discussions the Bell Laboratories, on behalf of the Western Electric Company, undertook to work out and get into operation a plan and procedures for controlling the ballistic quality of 105mm howitzer high explosive ammunition from a single loading line. The loading line selected was that at Joliet Arsenal. Work under the contract continued from July 1951 to October 1952. It was the intent that the procedures developed for this one loading line should serve as a pattern for extension by Ordnance to other 105mm lines and also to ammunition of other calibers and types.

The project itself was carried out at the Ordnance Ammunition Center by a team composed of a small group of Bell Laboratories engineers each with an Ordnance "opposite number." This team received the help and cooperation of various divisions of the Ordnance Corps as well as contractor personnel from component manufacturers. Such success as the project has attained is thus the result of the efforts of many people.

II. The Problem

It is important, at the outset, to understand that the objective of this project was confined to the control of those quality characteristics of the 105mm high explosive complete round which affect hitting the target. This was to be done within the limits attainable commercially with the present design of these rounds. The investigation was not concerned with a general system of process controls for all the types of defects for which Ordnance inspects a vendor's product. For example, it was not concerned with problems of assembly or mating of parts nor even with the operability of fuzes. Where, during the course of the work, it was apparent that production or testing methods could be improved, this was of course brought to the attention of the responsible Ordnance group.

Basically, then, the objective was to control production of the 105mm complete round such that when a relatively large number of these rounds are fired under the same conditions, the pattern of hits will have minimum dispersion. This dispersion must be minimum not only with rounds assembled from components, each type of which comes from a single manufacturer, but also when components of an individual type are supplied by several manufacturers.

III. Development

The orderly attack of this problem involved a series of rather distinct steps which, in retrospect, appear to be similar in many respects to those often followed in setting up a control procedure on a line within a plant.

A. Isolation of the Significant Causes of Ballistic Variation

The first step involved the identification and isolation of the significant causes of variation in muzzle velocity and range as well as the physical characteristics which largely contribute to these causes of variation. This was necessary to assure that all necessary controls were effected and that unnecessary controls were avoided.

Let us consider, for example, the variations in range that might be obtained under a given set of firing conditions. The distribution of hits might look something like that shown in Figure 2(A). Gun and atmospheric effects contribute to some of this spread. Variations in physical characteristics of the shell, powder and primer contribute most of the rest. Suppose now that in a complete round lot the rounds in one portion of the lot differed consistently from those in another portion in a characteristic such as the weight of propellant charge or the tightness of the rotating band on the shell. The effect would be to give two similar distributions as in Figure 2(B), one with a high average range and one lower. The net result is an increase in the over-all dispersion. Some other factor, like bourrelet diameter, might have a different type of effect. Differences in bourrelet diameter between portions of a lot could result in differences in initial yaw of the projectiles as they leave the muzzle. In general, the larger the yaw the greater the drag effect and the smaller the range. Increased variation in the initial yaw is reflected as increased range dispersion, so that the net result is again two different distributions, one portion of the lot giving a larger spread than the other and probably with a different average, as in Figure 2(C).

The effect of a change in weight of propelling charge or of a change in tightness of the rotating band is to change the range by virtue of a change caused in muzzle velocity. These are interior ballistic effects. The effect of the bourrelet change or of a surface finish change is on the flight of the shell and are exterior ballistic effects. The effect of a change in weight of shell is both interior and exterior, since the heavier shell generally will have the lower muzzle velocity but for any given muzzle velocity will travel further. Thus, these two effects of shell weight tend to counteract each other. At the low charges with low muzzle velocities and shorter ranges the interior ballistic effect dominates and the net result of an increase in shell weight is a decrease in range. At the high charges with high muzzle velocity and long ranges the shell has a longer flight through the air and the exterior ballistic effect dominates, resulting in an over-all increase in range with increase in shell weight.

There are many such causes of ballistic variations, some tending to shift the average of the distribution, others to widen the spread and still others which affect both the average and spread. These causes may appear and cause variations within individual complete round lots and also between lots. Some of these causes, of course, had larger effects than others. It was quickly established, however, that those characteristics which affect the ballistic properties of a round of 105mm howitzer ammunition are largely incorporated in its components prior to their arrival at the complete round loading line. Therefore, attention was mainly directed to significant variations of the component parts of the round.

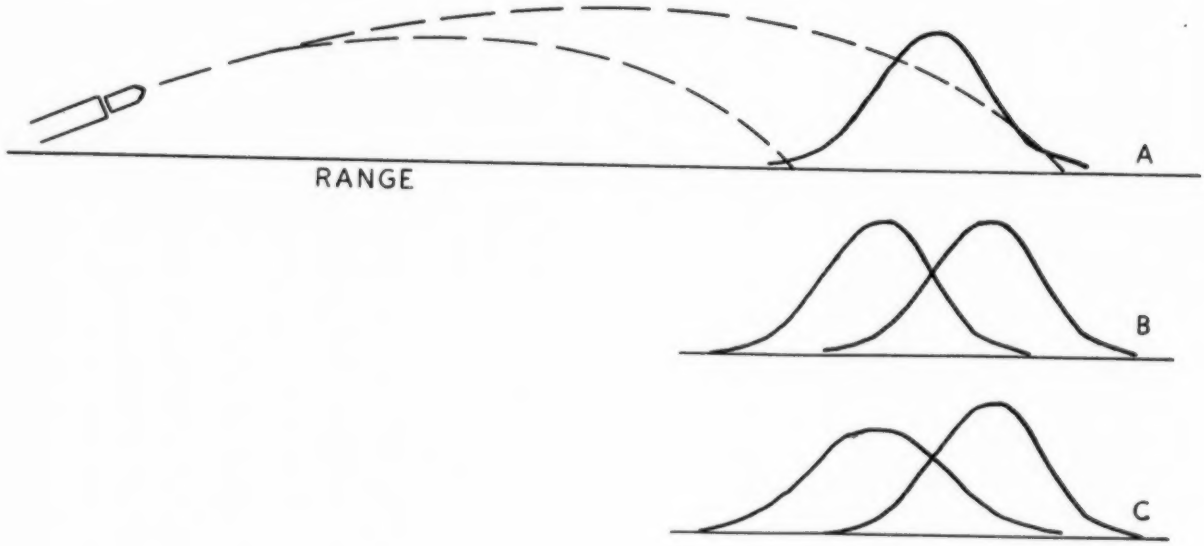


FIG. 2

To accomplish this phase of the work, both empirical and theoretical evaluations were made of the ballistic effects of variations in physical characteristics of components. Special tests were conducted where necessary and the personnel and reports of the Ordnance Corps were freely consulted. Ballistic acceptance test results, for components and complete rounds, were correlated with systematic differences in physical characteristics of similar type components supplied by several manufacturers.

As a result of these considerations the physical characteristics chosen for control were for:

Empty Shell

1. Weight and volume of cavity
2. Bourrelet diameter
3. Diameter and tightness of rotating band
4. Surface finish

Propellant Powder

1. Weight of bagged increments

Primers

1. Weight of black powder
2. Drying operation of the powder
3. Moisture content of the black powder

In the above, the primer and powder characteristics as well as the diameter and tightness of rotating band are included for their effect on muzzle velocity. Shell surface finish and bourrelet diameter are included for their effect on flight characteristics or exterior ballistic effect on range. The empty shell weight is included since it influences range both through its effect on muzzle velocity and on flight characteristics.

The volume of the shell cavity was found to be highly important in controlling empty shell weight. Cavity volume depends on one or two critical processing operations,--the piercing and drawing operations in forming the forging and also, in some cases, the subsequent cold-nosing operation. In addition, control of cavity volume itself is of importance in controlling the weight of the loaded shell. Therefore, even though there are no direct specification requirements on this characteristic, plans were made for including it as one of the important physical characteristics requiring control.

Thus, process controls were found to be necessary for characteristics of shell, propellant and primers. So far as ballistic quality control was concerned, there appeared to be no need for quality control procedures for the other components so long as design or construction changes in these items had no ballistic effect on range.

We cannot go into the factors influencing ballistic range in any detail here. Those wishing more information on this subject are referred to an interesting article by Colonel H. V. Wagner, on the basic causes influencing projectile behavior in flight. It may be found in the Sept.-Oct., 1952 issue of "Ordnance," a publication of the American Ordnance Association.(1)

B. Determination of Process Capabilities

The next step in the work was to determine the existing production process capabilities for those physical characteristics considered ballistically important. Process level and uniformity both had to be considered. Experience data were collected from a large number of periodic small samples. Although some data of this nature were obtained at the loading plant or proving ground by measurements made on the products of several manufacturers, for the most part they were obtained on the line at the component plant. This would not have been possible without the wholehearted cooperation of the component manufacturers.

For the most part, similar machines are used for corresponding operations at the various plants producing the same component. Their use, of course, was quite varied. The fact that they were all producing to the same specification limits had little connection to similarity of production tolerances. Few instances of statistical quality control were observed and, in general, 100% inspection to production tolerances was the rule.

With respect to weight of empty shell, data collected showed that each of several manufacturers had weight quite well controlled but at different levels. In all cases, the spread of weight among individual shell was much less than the specified tolerance, so that the specifications gave plenty of room for adjusting the average level of the manufacturing processes. In addition, if each of these distributions could be held fairly close to the center, practically all of the individual shell would be in Zone II when loaded. As you recall, this is the middle of the three shell weight zones I described earlier.

For the other shell characteristics and for primers, too, the several manufacturers were generally able to maintain their product well within the specification tolerances.

The case of weight of bagged increments of propellant powder was different, however. Here conformance to specified limits has involved serious manufacturing difficulties, particularly a need for frequent readjustment of measuring scales. Moreover the specification tolerances appeared tighter than necessary. In this instance a recommendation to liberalize the tolerances was made. The recommendation was accepted and the specification changed accordingly.

C. Establishment of Standards

The next step was to establish standards for the optimum value of each quality characteristic requiring control and for the degree of uniformity in successive components for these characteristics. This involved striking an economic balance between what the process capabilities had been found to be and what was needed in order to achieve optimum ballistics. Engineering judgment played an important part in striking this balance.

For some of the characteristics of components, notably the weight and moisture content of black powder in primers, all that was necessary was to hold individual items within the specified limits, with any reasonable shape of distribution. For weights of bagged increments of the propellant powder it was sufficient to have control within limits which

exceeded the existing specification limits. However, for the empty shell characteristics it was found necessary to control more closely than the specified limits. In any event, the objective was to aim continually at a stable distribution for the important characteristics, the distribution having the same desired average value and the same narrow dispersion for the components in successive portions of output from any one manufacturer. Moreover, it was important to have all manufacturers use the same goals in their process controls.

An additional consideration in the establishment of standards for weight and cavity volume of empty shell was the possibility of the elimination of "pre-zoning" of shell at the loading plant. In establishing process capabilities it was observed that, with a given manufacturing set-up for empty shell, the spread of the weight of loaded shell was about the width of a single weight zone. However, different manufacturers operated at different weight levels and the distribution for any single manufacturer might center almost anywhere in the three zone weight bands. Moreover, the distribution might shift from time to time with shell from a single manufacturer. Loaded shell falling in different weight zones could not be assembled in the same complete round lot.

As a result, production difficulties in loading occurred when loaded shell in more than one weight zone would be found on the line. Shell falling in zones other than that designated for the complete round lot being assembled had to be removed from the line and assembled in another lot. These "zone-outs" often were so numerous that they could not be removed quickly enough and threatened the safety limits of the bay in which they were stored. To avoid these difficulties the loading plants resorted to weighing and "pre-zoning" empty shell before loading, thereby culling out those considered likely to fall out of the zone being run. Since cavity volumes were not used for "pre-zoning" this procedure was not entirely successful in preventing zone-outs, for light shell were usually associated with large cavities, which partially compensated for the lighter steel after loading with explosive filler.

During the course of the project, it was demonstrated that the processes affecting weight and cavity volume of empty shell could be so controlled that the weight average of loaded shell would fall very close to the center of the Zone II band. In this event only a small proportion of loaded shell would fall outside the Zone II weight limits. Accordingly, the standards for these characteristics were established to accomplish this and thus eliminate the necessity for the pre-zoning weighing operation.

With respect to surface finish, it was considered that uniformity of finish from shell to shell and not the degree of smoothness as such was of first importance ballistically. Although all manufacturers maintained their surface finish below the specified maximum, their levels of average finish were very different. It was possible that finer finishes than necessary were being obtained with additional expense. For this characteristic both minimum and maximum limits were established to obtain the desired control.

In establishing the standards for bourrelet diameters, consideration was given to the fact that, the smaller the bourrelet diameter, the larger the variation in initial yaw and hence the larger the dispersion of hits. Considering the process capabilities, therefore, the desired

level was set as close to the upper limit as possible without causing any appreciable percentage of rejections for oversize bourrelet.

For band tightness the standard level of control, for ballistic reasons, was placed at the smallest clearance attainable economically by the several manufacturers. Band diameter was centered about their common average.

As described previously, the specification tolerances for weight of bagged propellant increments were found to be more stringent than necessary. The standards were set to more liberal tolerances which now have been adopted by Ordnance in the specifications.

The standards selected for the weight and moisture content of the black powder in the primer were based on the specification values. The standards selected for the drying operation in the ovens were based on the demonstrated process capabilities.

D. Quality Control Procedures

The next step required the development of plans and procedures for assisting the manufacturers of components in maintaining conformance to the established standards, and which would provide Ordnance with adequate evidence of such conformance. It was important that these procedures be simple, that they call for as little inspection as possible and that provisions be made for varying the amount of inspection to accord with the degree of control shown by the processes.

The quality control plan which was finally worked out consisted of a set of sixteen Quality Control Procedures. Figure 3 shows, in outline form, the nature and interrelation of these procedures. It is seen that they fall in three catagories:

(a) Process Control Procedures
(for use at component parts manufacturing plants)

(b) Control Verification Procedures
(for use at the complete round loading plant)

(c) Proving Ground Procedures

1. Process Control Procedures

The Process Control Procedures cover the characteristics which have already been described as contributing in an important way to ballistic dispersion. They each describe where and when the samples should be taken and the number, size and selection of samples. They also describe the precision and method of measurement as well as the method for recording and computing data. The central lines and control limits for $\bar{X}$ and R charts are given as well as suggestions for the use of the control charts during production, including interpretation of the charts and some suggestions for process investigations. They provide criteria for reduced sampling and request that the process control data sheets and control charts be available for review by authorized Ordnance Acceptance Inspection personnel.

BALLISTIC QUALITY CONTROL
Shell, 3cm, Fixed, HE, M1
for 105 mm How M2A1 & M4
14 AUG. 1952

LAYOUT OF QUALITY CONTROL PROCEDURES (QCP'S)

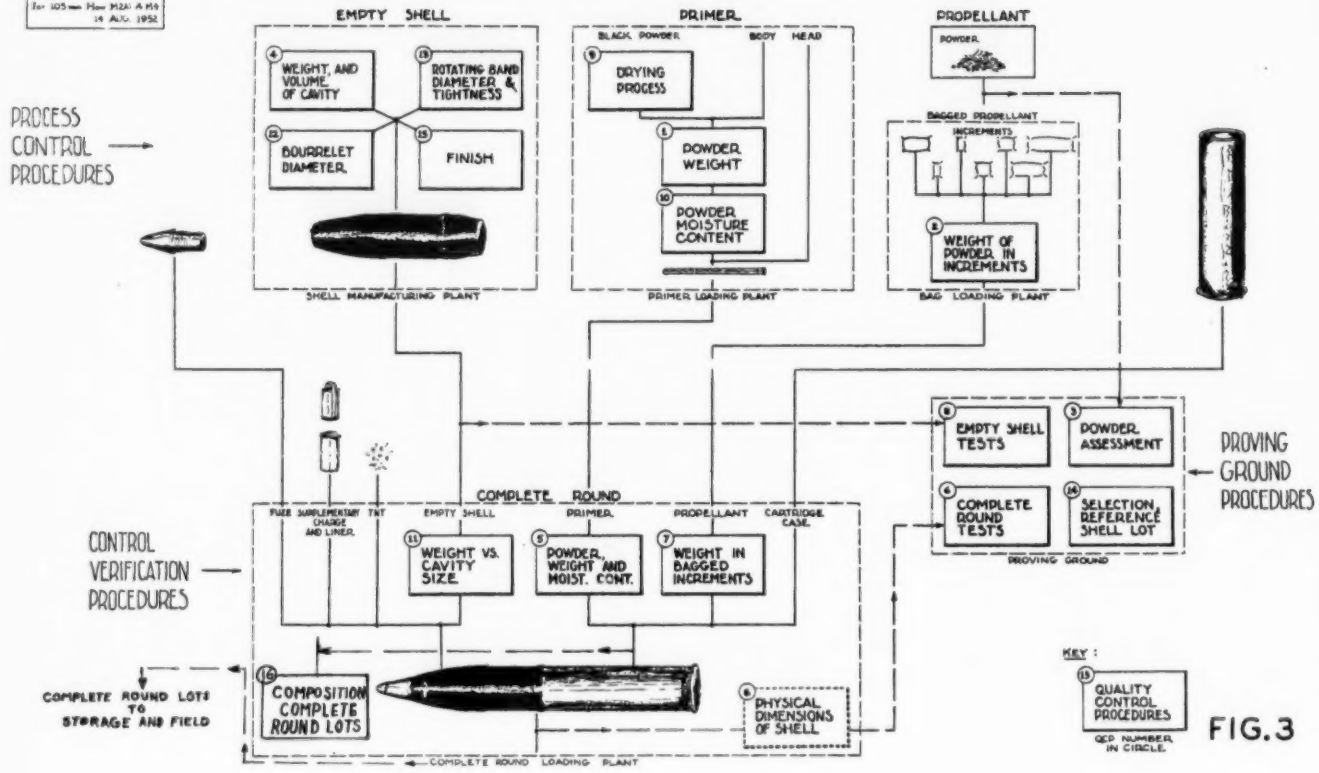


FIG. 3

These Procedures also contain an additional and highly important feature. They set up a rating plan whereby a lot will receive one of several ratings (AA, A or C) with respect to each of the characteristics under consideration. These ratings are based upon the control results obtained for the individual component lot. The rating information is placed on the lot data card by the Ordnance inspector at the plant and is thus made available to the complete round loading plant as well as to the Quality Control unit at OAC. Under this plan lots having an "A" rating or better may be combined in making up a single complete round lot. The rating plan also assists the loading plant in knowing when it can dispense with pre-zoning of shell, for shell lots rated "AA" or "A" on weight and cavity ordinarily do not require pre-zoning. In addition the rating plan serves another purpose--it gives a merit rating to the output of each manufacturer. Rating plans provide an incentive for good work and manufacturers generally are proud of a high rating when working in competition with others.

2. Control Verification Procedures

The Process Control Procedures were designed for use at the component manufacturing plant. Control Verification Procedures were developed for use at the complete round loading plant. As remarked earlier, there were no processes at the loading plant which contributed significantly to muzzle velocity or large variations. Ballistic quality control at the loading plant is then essentially concerned with the verification of the existence of adequate prior control of important component characteristics by the component manufacturer. This verification involves the evidence represented by the rating information on the data card and, in addition the application of Control Verification Procedures. The latter involve the measurement of a relatively few small samples to verify that no impairment of quality has occurred in storage, handling or shipment and that gages and test equipment have not gone out of calibration at some manufacturing plant. These procedures provide assurance and help to avoid the assembling of faulty material.

An additional Procedure (QCP-16 in Figure 3) specifies which lots of components can be combined to make up an individual complete round lot. This procedure also gives a method for dealing with lots of components which do not so qualify, because of failure to meet control criteria.

3. Proving Ground Procedures

Four Procedures were prepared for use at the Proving Ground. One of these (QCP-3) presents a standard method for the charge weight assessment of a powder lot. Another (QCP-14) presents a method for selecting reference or calibration lots of shell of as uniform a quality as is practicable. There are also two Procedures (QCP-6 and QCP-8) which give ballistic test programs to be followed at the Proving Ground for complete rounds and for empty shell respectively. These Procedures call for measurement of the ballistic samples for some physical characteristics which are not covered by the other Procedures, with a view to detecting new sources, if any, of ballistic dispersion under future conditions of manufacture. The chief purpose of these latter two Procedures, however, is to serve as an over-all check on the ballistic uniformity of the 105mm Howitzer ammunition.

E. Putting the Procedures Into Effect

As a final step in the development of the program, these procedures had to be put into operation on the complete round loading line. The initial loading line selected, that at Joliet Arsenal, received components from a number of manufacturing plants for assembly, so that it was necessary to install some of the procedures at most of these other plants. Several manufacturers supplied the Arsenal with empty shell and one of the objectives of the procedures was to obtain uniformity among their respective products by having them all work toward the same standard levels.

An important feature of the installation of these procedures by the shell suppliers is that it was entirely voluntary on their part. It was not considered advisable at that time to change drawings or specifications in accordance with the new limits in the procedures. This policy was followed for several reasons. Little experience was available on the length of time necessary to achieve controlled processes for the various characteristics. Changes of drawings would undoubtedly have meant rejection of material which was satisfactory by present standard. Moreover, the bulk of the tolerances used in the present drawings have been proved by experience to be easily attainable; in the event of an expansion of the base of production we might be faced with the problem of new inexperienced suppliers unable to meet tighter limits. Finally, this policy was followed in the conviction that process quality control must be understood, used and believed in by the producer if it is to be effective, - it cannot be obtained by mere legislation.

In our discussions with shell suppliers, it was made clear at the outset that there would be no rejections of shell meeting the present drawings and specifications. The very real advantages to the troops in the field to be obtained by the supply of ballistically uniform ammunition were described. The accumulation of data under this program was shown, supporting our conclusion that this ballistically uniform ammunition could be obtained by control of a few critical characteristics within present process capabilities. The very real benefits accruing to the manufacturer himself by use of process quality control were explained.

In view of the differences in types and use of machines among the various shell plants and the prodigious task of providing machine process control procedures for each operation within each plant affecting one of the critical characteristics, the procedures for use at the shell plant called for sampling at the end of the line. The intent was for these procedures to show the degree of process control exercised by the manufacturer. Where lack of control was evident, it was then expected that the manufacturer would go back to his process to find and correct the cause. An example of how this has successfully operated in some of the shell plants will be shown later.

To illustrate some of the problems encountered during the initial installation of the procedures at the shell plants consider the case of one manufacturer, whom we shall call "A". "A" was quite willing to obtain the data called for in the procedures but balked at following the provision in the procedure which said that when lack of control was evident he was to go back to his process to find and correct the cause. The plant was characterized by tight production tolerances, 100% go-no go gaging and a generally good quality, insofar as dispersion of his

product was concerned. "A" estimated the cost of collecting the data and felt that this cost should be borne by Ordnance since he felt the data would be more useful to Ordnance than to him.

A controlled product would result in subsequent savings to Ordnance since it would simplify loading and handling problems, save ammunition and increase accuracy in the field. Although these savings would exceed the cost quoted by "A", mere data collection does not improve or even maintain quality of process. Unless the procedures were used to take corrective action where indicated, they would not be used by the manufacturer to obtain process control. "A" was interested, however, in any suggestions for improving his method of process inspection through process quality control at the machines, if it could be demonstrated that any reduction in inspection costs would result. It was agreed then that a process quality control procedure would be developed on a trial basis at one of the finish grind machines for bourrelet diameter. At this centerless grinder, one of many in a battery, statistical quality control would be developed in parallel with the existing inspection practice, which had one man inspecting the output of each machine.

Personnel from the Quality Control unit at OAC worked together with the plant in developing the procedure for use at the machine. A number of improvements in operation of the grinder resulted. Extension of the machine process control to the other machines in the battery resulted in the elimination of half of the inspectors. The plant put the inspectors no longer needed on these machines to obtain process capabilities on other operations and withdrew its request for compensation for the "cost of taking data."

Another feature which has been helpful in the installation of the process control procedures is a Quality Control Performance Report which is used by the Ordnance Inspector at the manufacturing plant. This is a form which the Ordnance Inspector prepares whenever two successive lots receive a "C" rating for any physical characteristic covered by a procedure. In addition to listing the measurements made on that characteristic for the lots in question, the inspector lists the causes believed by the contractor to be responsible for the lack of control and also the corrective action taken. This report is sent to the Ordnance District and to the Quality Control unit at OAC. It has been most helpful in pointing up common areas of difficulty where they exist and also has permitted removal of sources of trouble at one plant which have been successfully overcome at other plants.

The success of the project requires the production of uniform shell and not a lot of paper. The occasional use of this form by the Ordnance Inspector and his notation on the data card reporting the grade received by the corresponding component lot are the only records which are required from the component plant in connection with this quality control plan.

IV. Results to Date

The installation of the quality control procedures for the loading line at Joliet Arsenal has now been extended to other loading plants as well. This involves, then, all regular producers of 105mm shell. All shell suppliers that have been contacted have indicated their desire to cooperate in the program. Only one 105mm shell manufacturer has yet to start rating his shell and he is expected to start soon.

As of March 1, 1953, 73% of the rated lots are rated "AA" or "A" on weight and cavity volume and do not need to be pre-zoned. This and similar percentages for other characteristics are listed in the table below.

Characteristic	Rated Lots with an "A" Grade or Better
Weight & Cavity Volume	73%
Front & Rear Bourrelet Diameter	74%
Band Diameter	79%
Band Tightness	90%
Surface Finish	48%

We have been in the process of changing the surface finish measurement to be taken before paint rather than after paint and it is expected that this characteristic will improve.

The procedures have been in effect at the different plants for varying periods of time. As an example of the success of their operation, consider their operation at the Kelsey-Hayes Wheel Company, which has kindly consented to our presentation of their experience. This company was among the first to install the quality control procedures. The history of their lot ratings is given in Figure 4. The steady improvement and achievement of control with respect to the desired standards is apparent. This was achieved by the plant setting up their own machine process controls for operations highlighted by the operation of the QCP's. Often these were the QCP's themselves brought back to the machine with slight modification. During the six-month period following installation of the control procedures the overall scrap was reduced to a third and rework to less than a fifth of that experienced at the start of the period. Together with this reduction the plant noted a smoother flow of material through the production line. At this plant interest in the success of the quality control program was evident throughout the chain of command. This included the machine operators as well, for as scrap and rework lessened, bonuses were increased.

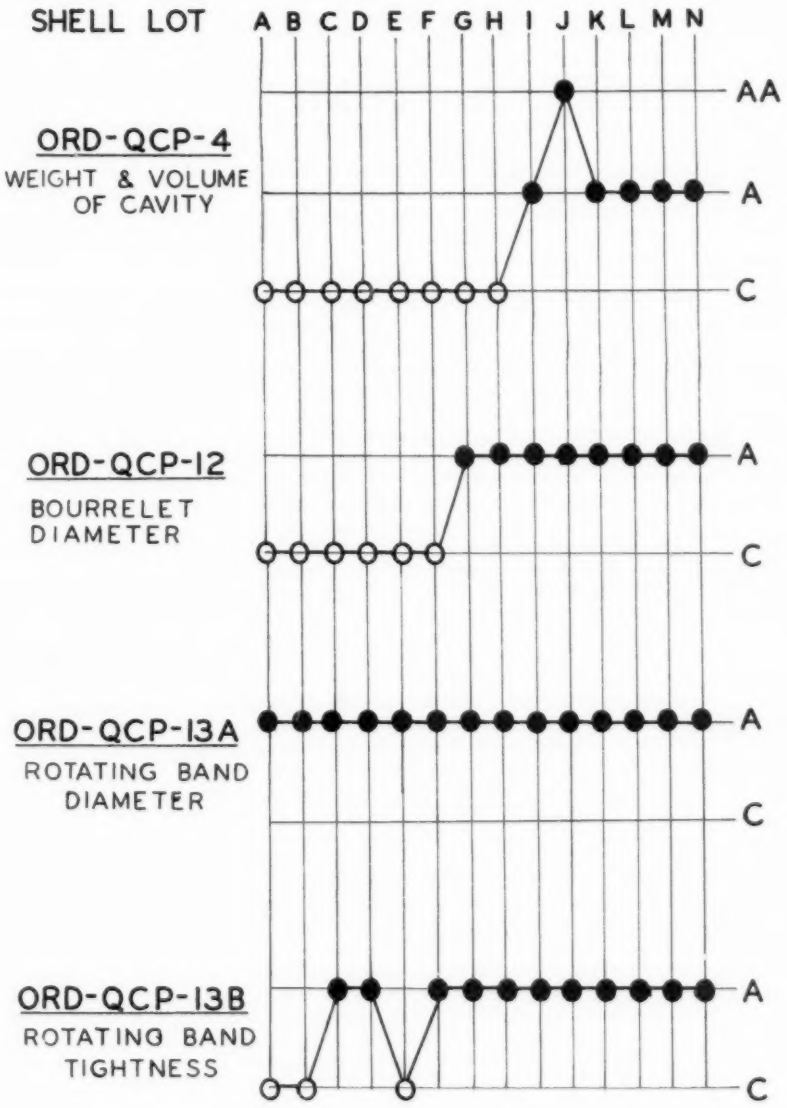
Since the initial installation of these procedures, many requests have been received for assistance in setting up process controls on other operations within the plants and on other production items. The net result has been an increased awareness of the benefits derived from knowing the capabilities of various processes and from quality control in general.

With reference to the matter of weight prezoning of empty shell at the loading plant, we are pleased to report that at one plant there has been practically no pre-zoning for about a year, with a negligibly small number of zone-outs after loading.

Perhaps the most important result of the ballistic quality control program relates to its progress in accomplishing the main objective of the program--to obtain ballistically uniform ammunition. This has been demonstrated in the successful production of very much larger lots of complete rounds having uniform ballistic characteristics.

Specifications have placed restrictions on the size and composition of complete round lots because of differences that have been known to

KELSEY HAYES WHEEL CO.



EACH SHELL LOT REPRESENTS ABOUT ONE WEEK OF PRODUCTION.

FIG. 4

exist in some cases among lots of a single type of component from a single manufacturer, as well as among similar lots from different manufacturers. Other factors, such as the necessity of weight-zoning, have often further restricted the lot size.

Regular manufacture of howitzer ammunition in very much larger lots has obvious advantages and is a step towards the ultimate goal of no ballistic differences among lots. The segregation of lots complicates the ammunition problem all the way from the loading plant to the artilleryman in the field. Storage, handling, shipment and use are all simpler with fewer lots. When an artilleryman begins firing ammunition bearing a different lot number, he must in general fire a series of spotting rounds. The resulting delay in getting on the target may have serious combat consequences as well as resulting in a very substantial expenditure of ammunition.

It was demonstrated, during the course of work on this project, that certain groups of complete round lots could each have been treated as a single large complete round lot without increasing dispersion of muzzle velocity or range. The components assembled in these lot groups had met the criteria of the several Quality Control Procedures. Application of these procedures has now resulted in the successful production of a number of large lots whose size is over five times that of lots produced prior to the introduction of the Quality Control Procedures. The dispersions of muzzle velocity and range within these large lots were fully as low as for the very much smaller lots produced previously.

Throughout the program any change which would reduce production in any way at all was scrupulously avoided. On the contrary where these procedures have been applied production flow has become smoother and more good pieces are made.

The success of this program was indeed the result of the coordinated efforts of an industry-Ordnance team. Through the excellent cooperation of the various manufacturing plants it has been possible to achieve major benefits, not only those derived from the use of process quality control within the plants but also to the using services, from both a logistic standpoint and in actual firing during combat.

Literature Cited

(1) Wagner, Col. A.V., "The Projectile in Flight," Ordnance Vol. XXXVII, No. 194, Sept-Oct 1952, ps. 339-345.

NEW DEVELOPMENTS IN AIR GAGING

F. Meyer, Jr.
The Taft-Peirce Manufacturing Company

With the ever increasing demand for greater precision in the manufacture of component parts for many of the products that are now considered necessities in our present mode of living (airplanes, automobiles, electric refrigerators, etc.) there is constant need to develope equipment for manufacturing parts to closer tolerances and provide means of accurately checking the machined parts. While the demands are increasing for machining and gaging accuracy, the supply of skilled help available to perform the required production operations is decreasing; therefore, tool and gage engineers are constantly being called upon to devise methods and equipment that can be used by skilled operators to produce parts to a high degree of accuracy.

The problem of the gage engineer is to provide measuring equipment that can be used by unskilled help, and give results that in the past were only achieved by skilled gage makers with many years of practical experience. In order to achieve this goal, it is necessary to build the required precision into the gaging equipment, so that it will be practically foolproof and free from 'human error' even when it is used by operators of limited skill.

It was this need that presented the opportunity for the growth of the air gage as a production measuring tool.

During World War II, industry started to use the air gage for production inspection. Originally, this type of gage was only considered for the measuring of internal diameters whenever a high degree of accuracy was required. As the many advantages of the air gage became apparent, its application became more diversified, and it is now used for checking numerous dimensional relationships such as concentricity or alignment between two diameters, squareness bore to face, center distance between bores, angle of taper, etc. Quite often, several of these dimensional characteristics are checked simultaneously.

When the air gage is used for one of these measuring problems, the setup is complicated by the fact that there is always more than one variable involved, since readings are not only affected by variation of the above factors, but also by variations in bore diameter. The total effect of both variables must be considered in arriving at the final result. It is necessary, therefore, to use at least two air circuits and a like number of standard air indicators to arrive at the desired result. The operator frequently has to take readings from two air indicators, then add or subtract the respective readings before the final answer can be obtained. This computation is required when checking tapered parts, center distances between holes and clearances between mating parts. When checking concentricity between diameters or squareness of bore to face, it is usually not necessary to obtain results from separate indicators and then compute an answer, but due to the permissible variation in diameter involved, position of the tolerance range on the scale will change if diameter varies from its mean value.

If the standard type air gage circuit and indicator is applied to these measuring problems, more responsibility is placed on the operator

and the human element has not been entirely eliminated.

Computing Indicator

Recently, a new type of differential air indicator was developed which overcomes this objection. With this indicator, two air circuits are applied to opposing chambers of the differential device. The gaging nozzles are so placed in the gage member, that the effect from any variation of diameter will be applied on both sides of the gaging equation, thereby nullifying its effect on the reading obtained on the dial.

Thus, it is possible to obtain readings for these dimensional characteristics on a single dial with a single pointer. Since the indicating device receives the results of two gaging circuits and makes an automatic computation of the resultant effect, it is called a "COMPUTING INDICATOR".

Illustrations of production applications of this computing indicator are shown in Figures 1 to 4.

Figure 1 shows an installation for checking the skirt section of a piston for diameter and taper.

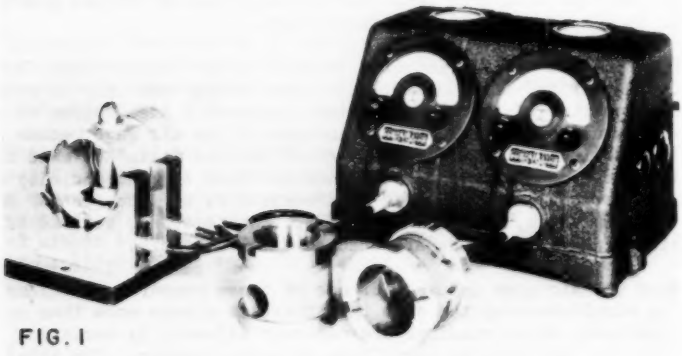


FIG. 1

The skirt diameter at the bottom end of the piston is checked by a standard air circuit indicator, while the taper of the skirt section is shown by a computing indicator.

There are no unusual construction features required in the gaging member, the design being exactly the same as when it is used with standard air indicators.

Figure 2 is a schematic layout of the gaging member shown pictorially in Figure 1.

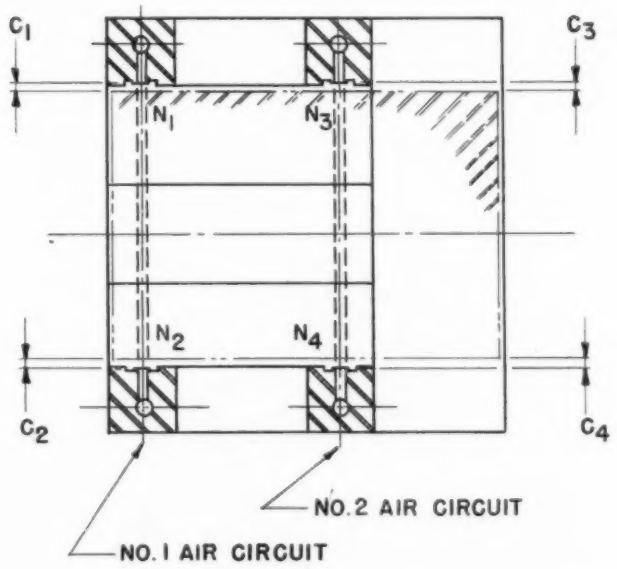


FIG. 2

Circuit Number 1 is connected to a standard air indicator which will show the diameter at the large end of the piston. In addition, a by-pass from this circuit is connected to one of the differential chambers of the computing indicator. Circuit Number 2 is connected to the other chamber of the computing indicator. If the diameter being checked by nozzles N_1 and N_2 should vary from the mean value, then the sum of the clearance $(C_1 + C_2)$ will also vary from the mean amount and the diameter will be shown on the standard air indicator. The effect of the clearances C_3 and C_4, as checked by nozzles N_3 and N_4, will be imposed on the other chamber of the computing indicator. If the angle of taper is at the mean value, but the diameters at (N_1, N_2) and (N_3, N_4) are greater or less than mean, then $(C_1 + C_2)$ and $(C_3 + C_4)$ will be increased or decreased by like amounts.

Since $(C_1 + C_2)$ and $(C_3 + C_4)$ are being applied in opposing circuits to the differential chambers of the computing indicator, and if their sums are increased or decreased by equal amounts, the variations on each side of the gaging equation will cancel each other and have no effect on

the resultant reading, and the angle of taper will be indicated as zero (mean) on the computing dial. From this description, it can be seen that diameter variation within the gaging range will not effect the indication of taper as shown by the computing gage. When the angle varies from the mean value and the summation of $(C_1 + C_2)$ does not equal $(C_3 + C_4)$, the amount of variation from zero will be indicated by the computing unit.

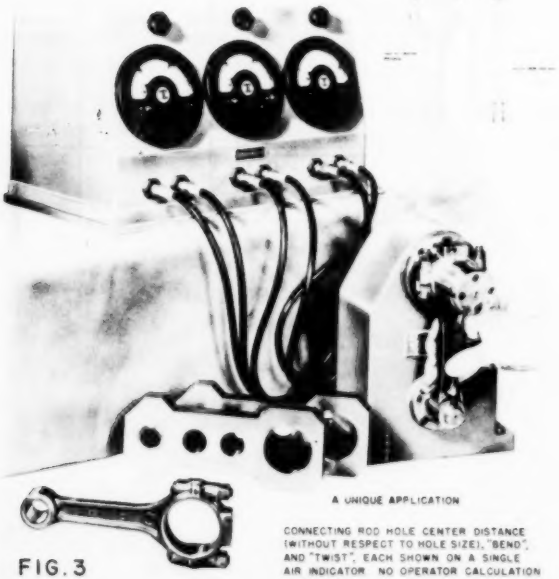


FIG. 3

A UNIQUE APPLICATION

CONNECTING ROD HOLE CENTER DISTANCE (WITHOUT RESPECT TO HOLE SIZE), "BEND", AND "TWIST", EACH SHOWN ON A SINGLE AIR INDICATOR. NO OPERATOR CALCULATION NECESSARY

The gage installation shown in Figure 3 is used for checking bend, twist and center distance of the bores in an automotive connecting rod. A total of six air circuits are employed and their gaging results are shown on the dials of three computing indicators. The exact values of bend, twist and center distance are read directly regardless of variations in bore size. Furthermore, even though there is a permissible .004 variation in center distance, the guide surface of the air plugs can be made small enough and the nozzles recessed sufficiently so that the gaging fixture is made with fixed plugs, eliminating any moving parts in the gage member.

A schematic layout of the center distance gaging circuits is shown in Figure 4.

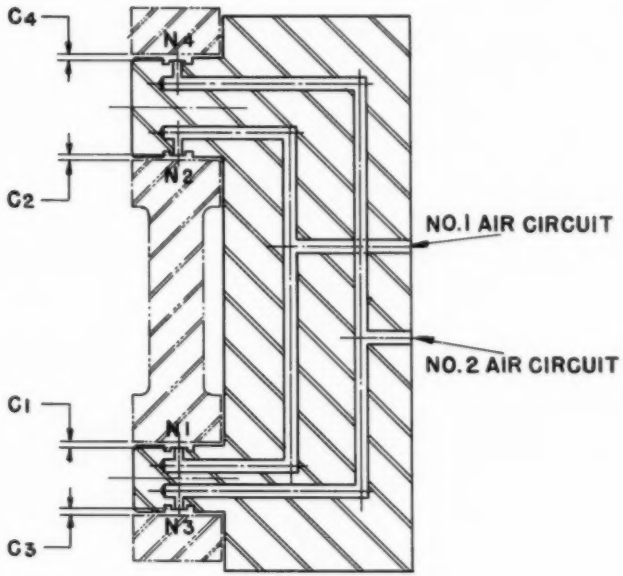


FIG. 4

It can readily be seen that since the algebraic principles involved are similar to those previously described for the checking of angle of taper; therefore, variation in bore diameters will not effect the gaging result shown on the computing dial. Bend and twist conditions are checked in a similar manner, through properly positioning of nozzles in the air plugs of the gaging fixture.

In addition to the illustrations discussed herein, there are numerous other gaging applications which can be satisfactorily checked with a computing air gage circuit. This type of equipment can be used to check bore squareness to face, concentricity between diameters, and the clearance or interference that will exist between parts which will be mated at final assembly.

Air Gage Members for Wide Tolerances

Another recent development in air gaging is the special contact device designed for checking wide tolerance ranges.

Contact type air gage members have been used for several years for

checking parts with surface finishes beyond 50 R. M. S. or wherever the section of the part to be gaged is of a limited area which can not be checked by having the air impinging directly against the part surface. These members are usually designed with a contact button mounted on a reed-type spring and positioned directly over an air nozzle. The contact button or buttons engage the part surface and indications of size will be obtained according to the resulting clearance between the reference or back face of the contact button and the surface of the air nozzle.

Original designs of the air contact gages were limited to a maximum gaging range of .008", the same as that of the direct jet type of gaging member. However, by design variations in the angle of approach between the reference face of the contact and the surface of the gaging nozzle, it is possible to check tolerance ranges up to .125".

The gaging members will be of a construction similar to the one illustrated in Figure 5.

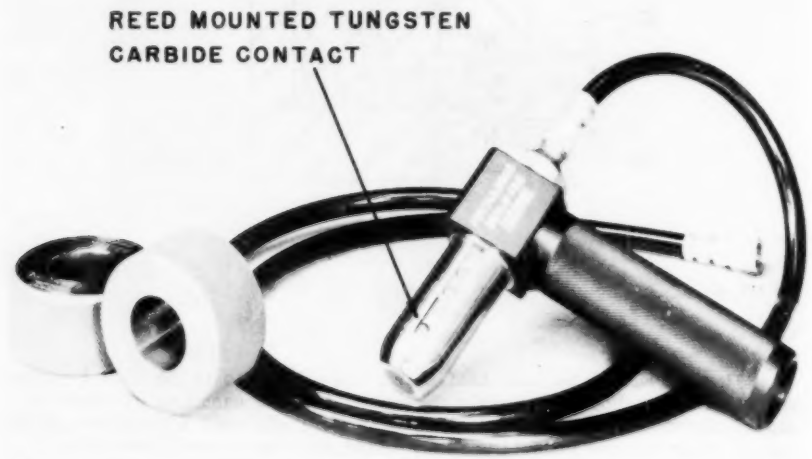


FIG.5

The single reed mounting of the contact can be followed for tolerances below .015". Beyond this figure, the contact should be mounted on parallel reed springs, or in some instances, on an anti-friction ball slide. The type of contact mounting will be dependent on the individual application involved.

Through the use of these special contact members for checking wide tolerances, the scope of air gaging has been greatly increased. In most instances, it is now possible to design suitable air gaging equipment to check any specific tolerance range from .0001" to .125".

Although extreme accuracy is not usually required when the tolerance range is quite large, it is still possible to reap the other advantages of air gaging such as - speed of indication; simplicity of gage design with no moving parts required to transmit gaging results from the member to the indicator; exact indications of size which are necessary for plotting statistical quality control charts; gaging in restricted areas which would not be accessible to mechanical devices; transmitting gaging results from the member to the indicator through a small flexible tube, thereby making it possible to check many dimensions simultaneously and reading the results on a panel containing several air indicators.

An excellent example of this type of application is the checking of the contour of the turbine blades. As many as thirty-nine dimensions can be checked at one time, and even though the degree of accuracy of the dimensions involved does not require air gage precision, it would be impractical and almost impossible to apply mechanical gaging to this application unless the inspection process was broken down into several operations.

The examples shown here cover only a small portion of the potential air gage applications. The use of this measuring means is constantly increasing with the requirements of industry for a versatile and rugged measuring tool that can be used by machine operators to keep a constant check on the machining process and establish suitable data for statistical quality control charts.

ARMY ORDNANCE GAGING PROBLEMS AND PRACTICES

M. L. Fruechtenicht
Frankford Arsenal Gage Department

In the United States the use of gaging in interchangeable manufacture dates back to the year of 1798 when Eli Whitney, the inventor of the cotton gin, received a contract to manufacture muskets. He realized that if he followed the European methods of custom fitting each part that he could not meet his contract. Using his truly American ingenuity he set out to manufacture muskets on what was a production basis at that time using specialized machinery and a gaging system to produce interchangeable parts. The Whitney musket as designed and produced could be assembled without the usual painstaking hand-fitting, and the parts of several muskets could be interchanged. Whitney achieved this interchangeability by strict adherence to self-established gage and measurement standards. It was not until some time later that definite gaging and measuring standards were established on a national basis.

Today in the production of ordnance materiel, large, medium, and small manufacturing enterprises have government contracts. The manufacturing and inspection methods are such that the various component parts which comprise a complete item of materiel can be manufactured by sub-contractors throughout the country and when furnished to the prime contractor for assembly, will fit into the assembly and function as intended. The Ordnance Corps is responsible for the production of the required materiel for the various using services within the Department of the Army.

The Ordnance Corps is organized on a mission center basis, that is, a certain installation is completely responsible for the development, product engineering and procurement of a particular type of ordnance materiel. For instance, Detroit Arsenal handles all matters pertaining to tank and automotive materiel, Watervliet Arsenal handles all cannon, Springfield Armory handles small arms weapons and Frankford Arsenal is a sort of hybrid in that the mission responsibility here includes small arms ammunition, fire control instruments, VT and MT fuzes, certain cartridge cases and shell as well as responsibility for coordinating the Ordnance Gage Program on a national basis. The term "fire control instruments" refers to precision optical and electronic equipment required in directing the fire of both mobile and stationary type weapons as opposed to equipment used for extinguishing fires.

The Ordnance Corps also operates a network of 14 district offices in the larger cities which handle the administration of contracts in their area and the acceptance inspection of materiel produced on those contracts. I imagine that the firms that many of you represent have Ordnance contracts that are currently being administered through a district office.

The gage program is organized on a similar basis. Each of the manufacturing arsenals or mission centers maintains a gage organization responsible for the design and supply of gages and inspection equipment

for their assigned type of materiel. Each arsenal maintains a gage laboratory for performing the dimensional inspection necessary for the acceptance of gages procured for their program. The gages are then issued by the mission arsenals to the district having responsibility for acceptance inspection. The district, in turn, issues the gages to their resident inspector at the plant producing the item.

Each of the district offices maintains their own gage laboratory and a small staff of gage engineers or technicians. The gage engineers are concerned primarily with assisting their resident inspectors and the contractors with their gaging problems. Where possible the districts discuss the gage problem and the scope of inspection, as outlined on the Standard Inspection Procedure, with the contractor at the time the contract is placed. The district offices, as a regular procedure, provide the contractor with a set of prints of ordnance gage drawings and gage lists, if they are available. The primary purpose for supplying these drawings is of course to provide the contractor with information as to what dimensional characteristics will be inspected so that he may set up his quality control program for process inspection along the same lines. Where a new item is involved and gage drawings are obviously not available, the gage engineer stands ready to assist in any way possible and if he can't answer your questions, he will refer you to the mission arsenal where I am sure you can get service.

Many contractors, of course, through their experience in gaging during World War II and through the quality control and gage programs they have established for their own product, do not require this type of assistance. However, it has been our experience that the high percentage of contractors that get into the ordnance production program do not have, or have not had a previous occasion to have, an organized gaging system. We have found that the district offices can be of considerable service to these contractors.

The district gage laboratories are established primarily for the purpose of maintaining surveillance inspection over final inspection gages in use in the various contractor's plants within the district. This function is one of the most important cogs in the ordnance production machine. We learned this through several rather unfortunate experiences at the out-set of World War II, however, since we had the nucleus of a district gage laboratory set-up established at that time, we were able to remedy the situation quickly. Each of these laboratories is currently maintaining between 10,000 and 20,000 gages under surveillance. When ordnance final inspection gages are not available, contractual arrangements usually provide that the contractor shall make his process gages available to the government inspector in order to accomplish acceptance inspection. When the contractor's gages are used for final acceptance, it is necessary that the district gage laboratory certify the dimensional correctness of the gages and this is usually accomplished by direct measurement in the district gage laboratory. The gage laboratories are also responsible for maintaining the gages in the hands of the resident inspector up-to-date with the latest part drawings as authorized by the contract. The modification of gages which require correction is usually handled by the district through a contract with a local tool and gage shop.

A secondary purpose of the district gage laboratory is to assist the eight mission arsenals in handling the acceptance inspection load of new gages procured by the government. This arrangement has many benefits in that the district gage laboratory man can maintain close personal contact with the tool and gage shops and rejections can be discussed on a first hand basis.

Thus far, I have been referring to gages as a general category. However, to fully appreciate the entire program it appears desirable to discuss gages in their specific classifications; namely, inspection gages and manufacturing gages. Inspection gages are those that are used by resident inspectors of ordnance assigned to a contractor's plant for the acceptance of materiel. They are also used for final inspection in our government arsenals. These gages are generally designed and supplied by the Ordnance Corps. The second classification covers manufacturing or process gages and are those gages used by the contractor or the manufacturing division of an arsenal for process or screening inspection before presenting material to the ordnance inspector. These gages are designed and procured by the user. As a matter of general policy, ordnance final inspection gages are designed to provide for the inspection of the finished part to the part drawing in such a way that the contractor may have complete latitude in selection of his method of manufacture and dimensional control.

Ordnance inspection is generally conducted on a sampling basis in accordance with recognized quality control procedures with which most of you are extremely familiar. The inspection plan or the scope of gaging is laid down in what we term a Standard Inspection Procedure or SIP. The gage lists issued by our arsenal gage organizations are coordinated with the SIP's and, in fact, usually precede the SIP and serve as the scope of gaging until the formal SIP can be issued.

The scope of gaging or inspection required varies considerably with the type and end use of the item. Ordnance constantly strives to accomplish acceptance of materiel with the minimum scope of inspection that will insure that materiel will be safe for handling and will stand up under battle conditions. Correllary factors, of course, are the inspection cost as well as the conservation of manpower.

In line with this approach, wherever the type or function of materiel will allow, acceptance is accomplished based on a performance test or on an end item inspection basis. For instance, end item inspection is accomplished on certain fuzes and the only dimensional inspection by gaging is on the thread which serves to assemble the fuze to the projectile.

In the case of a complex item such as a tank which contains numerous complete unit type sub-assemblies, the scope of dimensional inspection for final acceptance is confined as much as possible to the inspection of those dimensions which affect interchangeability to insure that replacement can be accomplished in the field without need for further machining. Interchangeability gages which are designed to simulate the mating part or assembly are provided. Since certain problems of alignment, centrality, location, etc. in assembly are extremely difficult to cover by a dimensional specification, interchangeability gages are usually designed to insure interchangeable assembly rather

than to check exact dimensional specifications of the parts.

In those cases where individual parts of an assembly or sub-assembly are considered as replacement parts in the field, it is necessary to design and supply gages for the critical dimensions on these parts and to insure that the original as well as the spare part production meets these gages.

The cost of acceptance inspection as well as the cost of gages is, of course, of concern to all of us. As previously stated, the Standard Inspection Procedure outlines a minimum scope of inspection and classifies the dimensional features to be inspected as critical, major, minor and incidental. Critical defects are usually inspected 100 percent and consequently, automatic and semi-automatic gages are considered for these applications. The high initial cost of the gages is easily amortized by the subsequent saving in inspection time. Gages designed to inspect minor and incidental defects are usually of a simple type which will satisfactorily gage the component dimension.

The final acceptance inspection of ordnance materiel as conducted by the resident inspector is generally conducted on a sampling basis in accordance with recognized quality control procedures. Military Standard 105A "Sampling Procedures and Tables for Inspection by Attributes" is the guide used in the Department of Defense. Ordnance Inspection Handbook ORD-M608-9 on "Statistical Quality Control and Acceptance Sampling" is a guide used in quality control programs throughout the Ordnance Corps.

At this time I would like to review for you in a very general way some of the recent developments in the field of gage design. Several significant advances have been made in the application of optical gaging methods in the inspection of ordnance materiel. Alignment of various critical surfaces on tank hulls are now being checked optically. The purpose of the hull alignment gage is to achieve a rapid, efficient and economical check of the machining and welding of the hull suspension mounting surface. As the speed, steering qualities, wear and general performance of the vehicle is largely governed by the proper alignment of its suspension system this check can be considered critical. Optical alignment gages have enabled the inspector to check the relationship of all significant points in one simple operation. They further provide for the systematic charting of the dimensional accuracy of each hull for controlling and correcting machining. The manufacturer can be furnished specific information as to location and degree of inaccuracies for effecting corrective measures. It has been established that due to the peculiarities in each hull type and the method by which they are held for machining, certain adjustments from true alignment are required in the machine tools to produce correct results after removal from the machine. The complete inspection procedure is too lengthy to outline here, however, the device is essentially a series of telescopes and targets.

The determination of surface roughness on a production basis has long been a problem in ammunition production since the degree of surface roughness affects the flight characteristics of projectiles. The initial development of a pneumatic type comparator gage was conducted by ordnance. A contract was placed with a private facility for the

further development and perfection of the method and as a result a unit has been developed which is now in wide usage in ammunition inspection. The unit consists of a measuring head which has a flexible washer-type contact point mounted on a rubber backing to insure that the contact will conform to the general contour of the product and thus measure only the effect of surface roughness. A fixture has been developed and is now in use which provides for reading of surface roughness at three points simultaneously on the surface of a 75MM Recoilless Shell. This gage has proven to be fast, accurate and will repeat. A major problem exists in connection with establishing a definite correlation between the readings obtained on an air comparator type gage and those obtained on a direct reading tracer type unit. There is still considerable work to be done in this field and our engineers are now cooperating in several ASA projects on this subject. In summation, we feel that the air comparator type gage is a significant step forward in the production inspection of surface roughness.

The gaging of complex parts having a considerable number of critical dimensions has always presented a problem, especially when the location of a feature is dependent upon several dimensions and tolerances. Fixed gages must necessarily consider maximum metal conditions and thus the gage for such a feature may give results considerably in error if the part is made to minimum metal dimensions on the correllary dimensions. A great number of these problems have been solved by the application of optical projection methods in all of our ordnance arsenals. Several unique applications have been developed recently on the complex components of the Mechanical Time Fuze. Savings in inspection manpower will be considerable but the major advantage will be in the improved control of the product and consistent inspection results.

Another development which has created considerable interest is the Mobile Gage Laboratory which was designed by the Gage Engineering Division of Detroit Arsenal. The purpose of this unit is to provide a mobile temperature controlled gage inspection laboratory that can be shuttled between various contractor's plants in an ordnance district area to provide for "on the spot" surveillance inspection and thus eliminate the time consumed in returning gages to a central laboratory for surveillance which may also mean tieing up production. The use of this unit can also, to some extent, reduce the need for procurement of "back up" sets of gages which are normally needed to insure continuity of inspection operations while surveillance inspection is being conducted at a central laboratory.

I would also like to review our progress in the field of gage standardization. This is of particular interest in view of the economies that will be effected by this program over the ensuing years. The most common types of gages used for the final dimensional inspection of ordnance materiel are plain plug, plain ring, thread plug, thread ring, and adjustable snap gages. Up to and including World War II, it was the policy of the Ordnance Corps to prepare an individual drawing to represent these type gages when required for the inspection of a particular dimension. Since there were seven designing arsenals duplications in design occured due to the recurrence of identical dimensions on different items of materiel. Procurement was

accomplished by sending a print to each potential bidder which also was costly. During the past several years Military Standard catalogs have been prepared which provide complete specifications for widely used sizes of both go and not go plain plug and plain ring gages. Additional catalogs have recently been prepared for thread plug and thread ring gages for all standard series threads. Each gage listed has been assigned a stock number to tie in with the Defense Supply Management Agency program to establish one stock number for a commonly used item which is to be used by all services to effect maximum supply economy.

These Military Standards are already paying off in Ordnance in several phases of operation. The need for preparing an individual drawing of the gage has been eliminated since the designer now merely selects the proper stock number from the catalog and applies it to the gage list and his job is complete. This has saved considerable cost and manpower both at the arsenals and on contract design. Since copies of the Military Standards have been distributed to all normal suppliers of Ordnance gages it is now possible to process procurement by merely refering to the stock number and quantity required in invitations for bid. An overall ordnance stock control system for standard gages and measuring equipment has been established which serves to reduce the need for procurement as well as the overall ordnance inventory of these items.

Copies of Military Standards pertaining to gages are available through the Superintendent of Documents, Washington, D. C., at a nominal cost. It may be possible that some of you could derive certain benefits in using these standards in your own operations.

In summation, I would like to state that we feel that great strides have been made in the past decade in the art of inspection and gaging as applied in Ordnance. We like to feel that perhaps our program has had its impact on industry as a whole and that as the pendulum swings back your advances in the art of inspection through quality control will provide the basis for further reduction in the cost of acceptance inspection of Ordnance materiel.

ERRORS OF MEASUREMENT

H. J. Becker
American Steel & Wire Div. U. S. Steel Corp.

There are a variety of errors which can occur in the measurements of characteristics. These errors might be classified roughly as personal and impersonal. The personal type are those which occur when an individual does not read or record correctly the measurement which was actually revealed. This paper is not concerned with that general type of error but instead will attempt to describe in detail several methods for evaluating impersonal errors, those attributable directly to the measuring instruments.

This type of error can be further characterized as being of two kinds, namely; calibration, or accuracy and sensitivity. In calibration we are concerned with the observed measurement agreeing with or being the same as the true measurement. For example, if one were to place 500 lbs. of standard weights on a scale he would expect to read 500 lbs. on the dial or beam. If after 5 or ten such trials it was found that the average weight indicated on the scale was 500 lbs. then it is reasonable to assume that the scale is probably in calibration at that weight. The question of accuracy and sensitivity however, concerns itself with the variation of indicated results about the average. This characteristic might also be referred to as reproducibility - the ability to reproduce the same reading time after time under essentially the same conditions.

In a wire mill as in many other industries many physical tests are performed which are destructive. Since a test can only be made once on a particular sample it becomes necessary that the testing device remain in calibration as well as being accurate and sensitive. Such testing devices need calibration at regular intervals to verify their correctness. Tensile machines are no exception and they are generally checked with proving rings which have been calibrated by the Bureau of Standards in Washington. The calibration curve for the proving ring has two parallel lines between which the indicated load may be expected to vary even though a constant load is repeatedly applied. This variation in the comparative standard allows for a variation or difference to be calibrated between tensile testing machines. However, the proving ring calibration is a static test and may not reveal factors, such as bind or drag, which affect the actual test which is dynamic in nature. To check the performance under actual conditions obtained in testing it is our practice to cross check machines with repetitive tests on wire from one coil and applying the analysis of variance techniques quickly to reveal the presence of significant differences in calibration between the machines.

The following example illustrates the method and mathematics used in the analysis. Approximately fifty consecutive samples of wire were out from a coil and then thoroughly mixed. Forty-two samples were selected at random and placed in 6 piles of seven samples each. Each group of seven samples was broken on a different tensile testing machine by the same man and under essentially the same conditions. The observed readings are shown in Table I.

Table I

Machine Number	1	2	3	4	5	6
	1325	1342	1345	1315	1350	1345
	1355	1330	1310	1315	1330	1355
	1345	1324	1340	1330	1335	1350
	1340	1336	1330	1355	1345	1340
	1355	1342	1340	1315	1310	1340
	1355	1362	1310	1320	1360	1335
	1340	1338	1335	1310	1340	1360

The above data were coded for mathematical simplification by subtracting 1325 from each measurement. All of the subsequent arithmetic will be applied to these coded data that are shown in Table II.

Table II

Machine Number	1	2	3	4	5	6	Total Rows
	0	7	20	-10	25	20	62
	30	5	-15	-10	5	30	45
	20	-1	15	15	10	25	84
	15	11	5	30	20	15	96
	30	17	15	-10	-5	15	62
	30	37	-15	-5	35	10	92
Total	15	13	10	-15	15	35	73
Columns	140	89	35	-5	105	150	514

You will note that the total for each row and column has been shown as well as the grand total. Table III shows the squares of the values in Table II. Again the totals of the columns and the rows are shown.

Table III

Machine Number	1	2	3	4	5	6	Total Rows
	0	49	400	100	625	400	1574
	900	25	225	100	25	900	2175
	400	1	225	225	100	625	1576
	225	121	25	900	400	225	1896
	900	289	225	100	25	225	1764
	900	1369	225	25	1225	100	3844
Total	225	169	100	225	225	1225	2169
Columns	3550	2023	1425	1675	2625	3700	14998

If there was no variation in the wire that was tested and if there was no variation within each machine then each measurement on a machine would be the same. The difference between machines would of course readily reveal the errors of calibration. However, this just doesn't happen and the variations noted in Table I include the variation in the wire and the machines.

This glossary of symbols is used:

R =	Sum of the row	Number of degrees of freedom of Q is rc-1
C =	Sum of the column	Number of degrees of freedom of q_r is r-1
r =	Number of rows	Number of degrees of freedom of q_c is c-1
c =	Number of columns	Number of degrees of freedom of q_e is $(r-1) \cdot (c-1)$
T =	Sum of all values	F Table - Found in most statistics books.

To convert variability into variance divide by appropriate degrees of freedom:

ie; total variance = $\frac{Q}{rc-1}$

The following formulae are used with the data from Table II and III

Total Variation $Q = \Sigma X^2 - \frac{T^2}{cr}$

$$= 14998 - \frac{(514)^2}{(7)(6)} = 8707.62$$

Variability between rows

$$q_r = \frac{R_1^2}{c} + \frac{R_2^2}{c} \cdots \frac{R_r^2}{c} - \frac{T^2}{cr}$$

$$= 7629.8 - \frac{(514)^2}{(7)(6)} = 1339.42$$

Variability between columns

$$q_c = \frac{C_1^2}{r} + \frac{C_2^2}{r} \cdots \frac{Cr}{r} - \frac{T^2}{Cr}$$

$$= 8899.5 - \frac{(514)^2}{(7)(6)} = 2609.12$$

Residual variability or interaction

$$q_e = Q - q_r - q_c$$

$$= 8707.62 - 1339.42 - 2609.12$$

To determine if a significant difference exists between rows obtain

$$F_r = \frac{\frac{q_r}{r-1}}{\frac{q_e}{(r-1)\ (c-1)}} = \frac{(c-1)\ q_r}{q_e}$$

$$= \frac{5\ (1339.42)}{4859.08} = 1.378$$

Check against appropriate value found in F Table.

To determine if a significant difference exists between columns

$$F_c = \frac{\frac{q_c}{c-1}}{\frac{q_e}{(r-1)\ (c-1)}} = \frac{(r-1)\ q_c}{q_e}$$

$$= \frac{6\ (2609.12)}{4859.08} = 3.22$$

Check against appropriate value found in F Table.

F in Table at 5% for 5 and 30 degrees of freedom is 2.53.

The original hypothesis that is being tested is simply "There is no significant difference between testing machines." The F ratio which is used to test this hypothesis is found by dividing the between column (between machine) variance by the residual variance. If the result is greater than the value found in the F table under the appropriate degrees of freedom at the desired significance level than it is to be stated that on the basis of the evidence the hypothesis is rejected. In other words, "there is a significant difference." In the case in question the F value for 5 and 30 degrees of freedom at the 5% significance level is 2.53. This is smaller than the calculated F_c and is indicative of a significant difference between columns or in this case between testing machines. It is to be noted that a slight bind was found in tensile machine #4. After it was corrected and a similar test performed there was no longer a significant difference.

Since the question of differences between tensile machines is basically a comparison of variation between columns to that of within columns, one can proceed in the following manner and obtain the same conclusions.

Again

Q = Total variation as before

q_c = Variation between columns as before

$Q - q_c = q_r\ q_e$ = Variation within columns

Obtain $F_c = \frac{\frac{q_e}{c-1}}{\frac{Q - q_c}{c\ (r-1)}}$

Where q_c has c-1 degrees of freedom and $Q - q_c$ has c (r-1) degrees of freedom

For our example $Q = 8707.62$

$q_c = 2609.12$

$Q - q_c = 6098.50$

Then $F_c = \dfrac{\frac{2609.12}{6-1}}{\frac{6098.50}{6\,(7-1)}} = 3.08$ and hence significant as before at the 5% level

The above is often set out in tabular form as follows:

Source of Variation	Sum of Squares	Degrees Freedom	Mean Square
q_c Between Columns	2609.12	5	521.82
$Q - q_c$ Between Rows	6098.50	36	169.40
Q Total	8707.62	41	-

$$F_c = \frac{\text{Mean square between columns}}{\text{Mean square within columns}} = \frac{521.82}{169.40} = 3.08$$

The other type of error with which this paper is concerned is that of accuracy and sensitivity. It is necessary to know how well a machine, micrometer, balance or any other measuring device will reproduce itself. To establish these facts one uses an application of Analysis of Variance generally referred to as the five by five test. This test is based on the premise that if a series of measurements of a given characteristic are made on a single object, each successive measurement will, in general, differ from all the other measurements in the series provided the measuring instrument possesses sufficient sensitivity to indicate these differences.

To demonstrate, five different places of wire were weighed by five different individuals using the same balance. Table IV is a record of the values obtained by each of the operators on each of the units.

Table IV

	Trials				
Units	1	2	3	4	5
a	3.7879	3.7876	3.7879	3.7880	3.7878
b	4.1655	4.1655	4.1656	4.1657	4.1655
c	4.6624	4.6624	4.6624	4.6624	4.6623
d	5.3679	5.3678	5.3675	5.3678	5.3677
e	6.1897	6.1896	6.1898	6.1879	6.1896

The data were coded by multiplying by 10,000 and subtracting 37870 from all of the trials for the first unit, 41650 from the second unit, 46620 from the third unit, 53670 from the fourth unit and 61890 from the fifth unit. The coded results were entered in the Measurement Errors Analysis Form shown in Table (V) for which we are indebted to

Table V Measurement Errors-Analysis Form

Units		Trials I	II	III	IV	V	Unit Sums	Average $\bar{\bar{X}}_u$
a	X	1 9	18 6	35 9	52 10	69 8	86 42	98 8.4
	$X-\bar{\bar{X}}_u$	2 .6	19 2.4	36 .6	53 1.6	70 .4		
	$(X-\bar{\bar{X}}_u)^2$	3 .36	20 5.76	37 .36	54 2.56	71 .16	87 9.20	
b	X	4 5	21 5	38 6	55 7	72 5	88 28	99 5.6
	$X-\bar{\bar{X}}_u$	5 .6	22 .6	39 .4	56 1.4	73 .6		
	$(X-\bar{\bar{X}}_u)^2$	6 .36	23 .36	40 .16	57 1.96	74 .36	89 3.20	
c	X	7 4	24 4	41 4	58 4	75 .3	90 19	100 3.8
	$X-\bar{\bar{X}}_u$	8 .2	25 .2	42 .2	59 .2	76 .8		
	$(X-\bar{\bar{X}}_u)^2$	9 .04	26 .04	43 .04	60 .04	77 .64	91 .80	
d	X	10 9	27 8	44 5	61 8	78 7	92 37	101 7.4
	$X-\bar{\bar{X}}_u$	11 1.6	28 .6	45 2.4	62 .6	79 .4		
	$(X-\bar{\bar{X}}_u)^2$	12 2.56	29 .36	46 5.76	63 .36	80 .16	93 9.20	
e	X	13 7	30 6	47 8	64 7	81 6	94 34	102 6.8
	$X-\bar{\bar{X}}_u$	14 .2	31 .8	48 1.2	65 .2	82 .8		
	$(X-\bar{\bar{X}}_u)^2$	15 .04	32	49 1.44	66 .04	83 .64	95 2.80	
Trial Sums		16 34	33 29	50 32	67 36	84 29		103 32.0
Trial $\bar{X}_t$ Averages		17 6.8	34 5.8	51 6.4	68 7.2	85 5.8	96 32.0	104 6.4

σ_d' = Std. Error of Measurement = $\sqrt{\frac{\text{cell } 97}{20}} = \sqrt{\frac{25.20}{20}}$ = 105 1.12

Grand Total of squares 25.20 97

$3\sigma_d$ = Max. Allowable Variation in Trial Averages = 1.34 = 106 1.501

$3\sigma_d'$ = Total Error of Measurement = 4.503

Trial Average Control Limits	
$\bar{\bar{X}}$ 6.4	$\bar{\bar{X}}$ 6.4
$+3\sigma_d$ 1.5	$-3\sigma_d$ 1.5
Upper Limit 7.9	Lower Limit 4.9

Mr. Fred Trowbridge of Sentinel Radio Corp. For example, observer 1 weighed Unit A and obtained a value of 3.7879 grams which is entered in cell 1 as 9 after coding. Similarly subsequent values were recorded in cells 4, 7, 10 and 13. The values obtained by the other four observers were recorded in the parallel cells. Sums and averages are obtained in both directions and recorded in the appropriate cells. The sum of the trial averages is recorded in cell 96 and the sum of the unit averages is recorded in cell 103. These obviously must be the same and constitute a check on the arithmetic. The grand average is recorded in cell 104. Since variance is the square of the standard deviation, we obtain the difference between the observed value and the unit average for each trial and square the result. For example the difference between the value 9 in cell 1 and 8.4 in cell 98 is 0.6.which is entered in cell 2. This value is squared and entered in cell 3. The balance of the corresponding cells are similarly treated. The sum of the squares for unit A is recorded in cell 87 and similarly the sums for the remaining units are entered in the cells provided. The grand total of squares is recorded in cell 97. The standard deivation of the error of measurement is the square root of the quotient obtained by dividing the grand total of squares by the degrees of freedom, which in this example is 20. This value is recorded in cell 105. To determine whether a state of control exists between successive trials, the maximum allowable variation in trial averages is computed by multiplying the value found for the standard error of measurement by 1.34 and entereing this result in cell 106. This value when added to and subtracted from the grand average provides the control limits for averages of sample size five. These control limits are shown on the form and observation indicates that the trial averages are within the appropriate control limits. The total error of measurement is taken to be plus or minus three times the standard error of measurement which in this case is equal to 4.503. When this value is decoded a value of 0.00045 grams or 0.45 milligrams is obtained.

With this in mind it becomes quite obvious that when using this balance the sample weight selected should be such that the error of measurement is a negligible part of the total weight.

QUALITY CONTROL IN INTERNATIONAL HARVESTER FOUNDRIES

James M. Barrabee
International Harvester Company

When we consider the increased production and complexity of today's castings, it is little wonder that more and more foundries are turning to Quality Control procedures in their operations. A look at the various foundry departments shows many variables under some form of control - statistical or otherwise. In most cases where a continued attempt has been made to use the resultant figures to control casting quality, there has been a definite reduction in foundry scrap and rework. One company(1) has even gone so far as to furnish a complete operational history with each casting produced - "a casting with a pedigree."

The manner in which preliminary preparation is handled by the foundry and engineering departments can often make the difference in whether or not a casting can be produced satisfactorily - at a quality consistent with economy. Investigation should be made on the original blueprint (preferably a casting print) referring to the maintenance of dimensional characteristics as set by engineering and difficulty in core practice. Iron specifications, allowance and locating points for machining, etc., should also be closely analyzed. One of our plants is even planning to set up an experimental area where not only sample castings can be made, but also a study can be conducted as to the most efficient setup for molding practice. Control of raw materials can also be studied; for without consistency or definite knowledge of inconsistency, little can be done to isolate or make adjustments for some of the basic causes of casting defects. Items such as sand, pig iron, scrap, chaplets, lumber, shot, etc., should all be sampled to determine conformity to desired specifications. Additional checks can be made in layout of new or repaired pattern equipment, dryers, and chills even prior to the preparation of sample castings or cores.

We all realize that for years our foundries have kept records of casting production, scrap produced, molding efficiency, etc., but the effectiveness of these records was often lost in the maze of figures. The use of simple charts has provided a means by which these figures could be presented simply and effectively even to the most unskilled personnel. Trends can be readily observed and sudden fluctuations stand out clearly. In cases where only certain jobs were originally plotted, the tendency has been for the other operators themselves to ask for further installation on their particular operation. If we now take a look at some of the foundry departments, it can be seen that it is possible to add some form of effective chart control where previously numerical records were not used to best advantage.

Sand Preparation:

When operating any muller, especially where there is no mechanical charging, the use of standard weights in charging the various elements is essential. Too often the use of battered containers, "one shovelfull," or "one wheelbarrow of" in determining mixes, contributes immediately to inconsistencies in the amount of materials added to the muller. All elements should be weighed and added to a muller which has been cleaned at least once a shift. If this practice is not followed in the core room, the coating which clings to the side of the muller will absorb a substantial portion of the oil that is added.

It is possible to check the efficiency of charging by using a standard $\bar{X}R$ chart on which the weight of material charged either over or under the standard amount can be plotted at various intervals. This type of chart can quickly show whether the operator is following specifications or mixing at his own discretion.

Mulling time can also be measured, for it too can make the difference between good and bad mixes. A recording wattmeter can easily be attached to the motor driving the mulling wheels and it quickly indicates changes in motor load on the recording paper(Figure 1).

WATTMETER CHART FOR FACING SAND

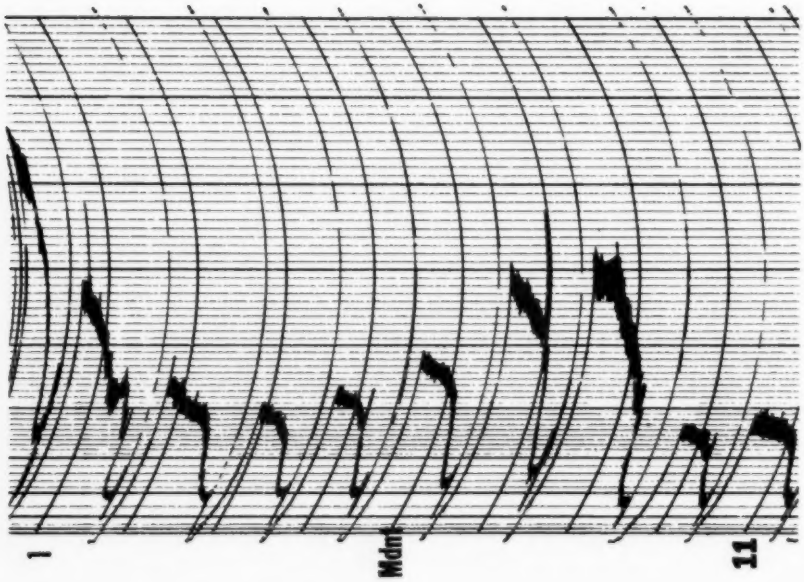


Fig. 1

The proper time for mixing each of the elements can be determined in preliminary tests and, from this point on, the meter merely records the time and loading of the motor throughout each of the mixing cycles. The buildup of green strength is also recorded on the paper and this can be calibrated so that the loaded muller will be emptied at both the proper time and at satisfactory strength. Thus, although not a true control chart, the running wattmeter record gives some evidence of efficiency of operation in a completely automatic manner.

Out in the core room itself, both green and dry checks can be made. Some foundries check cores visually in the green state to obtain information as quickly as possible, and then they sample the baked cores upon completion of cleaning and assembly operations for the overall quality picture. One floor molding foundry checks twenty-five each of the ten most troublesome cores on the molding floors each morning, and keeps a continuous P chart record of the per cent defective and type of defects found in each group. Another foundry uses a core rating system in which twenty cores twice a day are sampled from critical jobs. A rating of 10 is given for a perfect core, 8 for one with a minor defect, 5 for one with a major defect, and zero for a scrap core. A percentage value is then calculated by dividing the total points by a perfect score of 200 and compared with a previously determined standard. (Ex. 95% or 180 points out of a possible 200). Any deviation below the standard is recorded on a "Process Action" notice which is issued to the man at fault and his supervisor. A series of three such notices which are progressively colored white, yellow, and red are cause for warnings and finally disciplinary action to be taken against the employee.

One specific case illustrating the use of P charts for core checking occurred during an investigation of vent iron in some track idler cores. Poured metal often ran out through the vents and cracks in the core, and an average of 16.5% of the cores were found defective, mostly for cracks, unfilled vent holes, exposed wires, and dip surface flaws. Sample checking of the visual characteristics (Figure 2) was instituted daily on the

Fig.2

complete core assembly and over a period of time, a histogram of the completed charts showed a continual improvement down to the present 3% level (Figure 3).

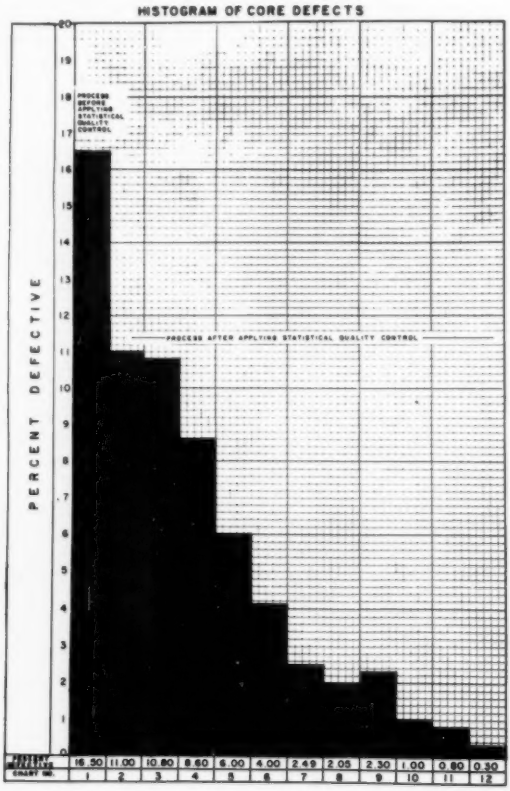

HISTOGRAM OF CORE DEFECTS
PERCENT DEFECTIVE
PROCESS BEFORE APPLYING STATISTICAL QUALITY CONTROL
PROCESS AFTER APPLYING STATISTICAL QUALITY CONTROL
16.50 11.00 10.80 8.60 6.00 4.00 2.49 2.05 2.30 1.00 0.80 0.30
CHART NO. 1 2 3 4 5 6 7 8 9 10 11 12

Fig. 3

A completely automatic moisture metering setup has been installed on one foundry sand system. A mercury tube sensitive to heat and two electrodes sensitive to moisture were placed in the weigh lorry carrying molding sand from the storage bins to the mullers. By measuring the temperature of the return sand and also its moisture, the system automatically calibrates and meters out a prescribed amount of water for that particular sand load into a reservoir above the muller. Once the sand has been dumped into the muller, the water is added immediately. In order to account for extreme variations in temperature and moisture due to nightly or week-end shut downs, a variable manual control of a base amount of water is possible, with the automatic controls taking over above a certain temperature and moisture.

A very short time after the system has been in operation each day, the sand reaches its normal variation pattern and from that point on no changes in base water are necessary and the metering is completely automatic. Not only has the control of moisture content been achieved, but the mulling time has been cut considerably taking quite a load off the operators who were continually pressed to meet the molding requirements.

Melt Department:

Aside from the usual analysis checks plotted during melt down, operational checks can also prove very effective in keeping the metal at proper temperature and analysis. Chill tests can be correlated against ferro-chromium or ferro-silicon additions where samples are taken from successive ladles and the resultant averages of chill readings can determine the amount of innoculation in the ladles. A new type of flowability test pad (Figure 4) made from cores is being used, where the

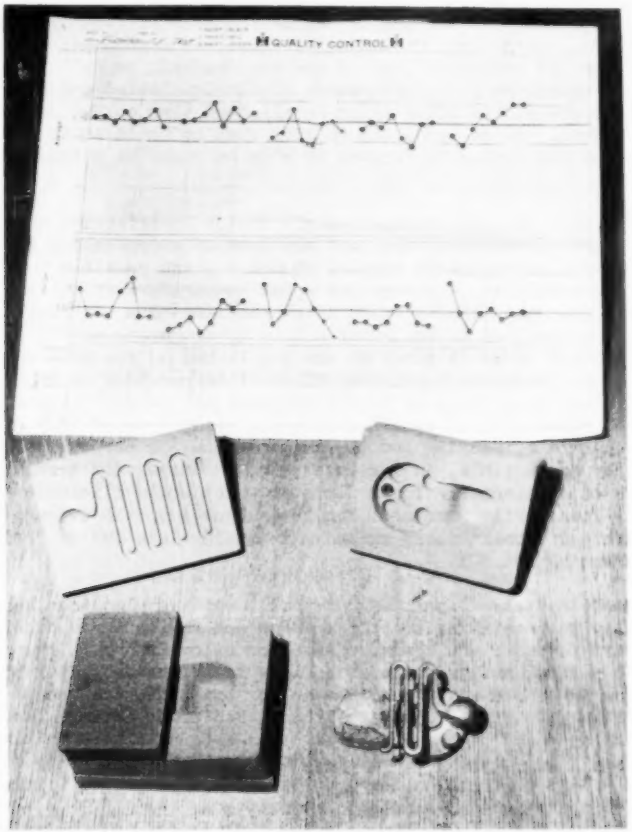


Fig. 4

length of flow in the spiral is very sensitive to changes in analysis and also in temperature.

Readings are plotted on an $\bar{X}R$ chart and they give a visual indication of this variation in temperature and analysis.

Another charging operation on which $\bar{X}R$ charts are applicable is on the cupola. Quite often a charging operator will attempt to level out one underweight charge by overcharging the following bucket, resulting in extreme variation in cupola operation throughout the heat. By setting charge standards and measuring the weight over or under the specified figure, similar to the muller charging operation, the variation can be easily measured and steps taken toward its correction.

A recent innovation has been the use of $\bar{X}R$ charts for job qualification on the pouring stations. A standard pouring time was set for each type of mold on the line based on defects and scrapped castings over a long period. Each hour the amount of time in 100ths of a minute over or under each mold standard time was measured and averaged on five molds and plotted on an $\bar{X}R$ chart. This gave an indication whether proper pouring practice, especially in keeping the sprue properly filled, was being observed by the iron pourers on the unit. Also, as mentioned, the charts were used for job qualification. A new man desiring to fill an iron pouring classification had his performance similarly plotted and either qualified or was disqualified after an allotted time on the basis of the $\bar{X}R$ chart record. He at least had to show that he could equal the record of the man he was trying to replace or else he would be automatically rejected.

"Cold" iron was the greatest single factor in defective casting rejection at one of our foundries; and the lack of superheating devices or a holding ladle necessitated tapping at the highest possible temperature. It was discovered that, although the spout temperature at the cupola was high enough for most jobs, present metal handling from the cupola to the molding floor resulted in too great a drop in temperature. Pyrometer readings plotted on an $\bar{X}R$ chart at the cupola and on the molding floor had proved the necessity for increased metal temperature as well as improved handling procedures to assure constant hot metal. (Figure 5). Subsequent investigations resulted in a change in type of coke used, a revised schedule of pouring and tapping to allow the cupola to run as continuously as possible, an increase in coke charge, and a relining of all ladles to a reasonable inner diameter which would eliminate the large heat loss during pouring. Figure 6 shows that the average cupola temperature was substantially raised and casting loss due to "cold" metal was cut almost 50%.

A recent article[2] pointed out that incentive plans can be used successfully in benefiting both management and the employee, where electric arc furnaces are used for steel or alloy iron melting. Conformity to desired analysis specifications is the prime factor as measured on $\bar{X}R$ charts and the melter-helper teams receive bonuses according to quality first, quantity second, and power savings last. The statistical approach was used on this bonus type plan because time study was not feasible. Melters and helpers are more alert, looking for ways to improve their performance through better controls and a friendly spirit of cooperation and also keen competition.

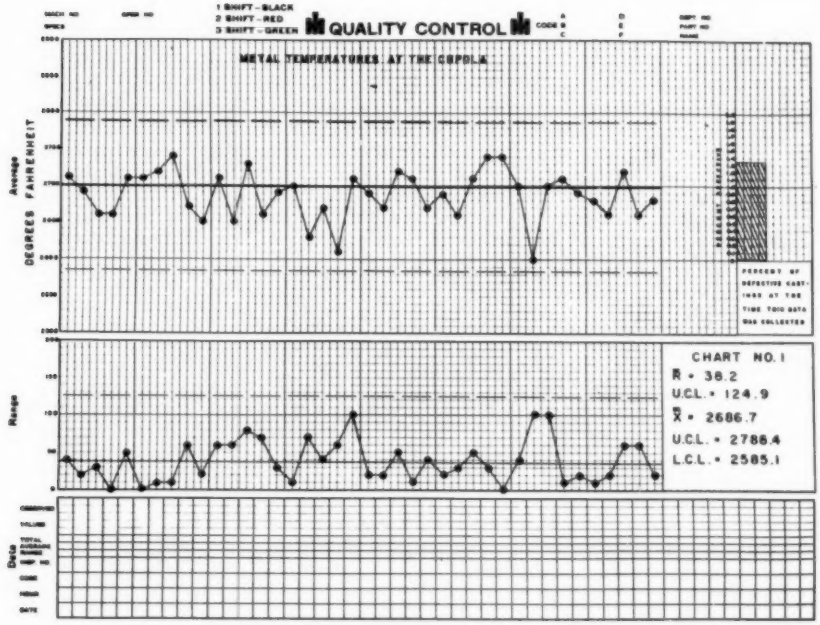

1 SHIFT - BLACK
2 SHIFT - RED
3 SHIFT - GREEN
QUALITY CONTROL
METAL TEMPERATURES AT THE CUPOLA
Average
DEGREES FAHRENHEIT
Range
Data
CHART NO. 1
R̄ = 38.2
U.C.L. = 124.9
X̄ = 2686.7
U.C.L. = 2786.4
L.C.L. = 2585.1

Fig. 5

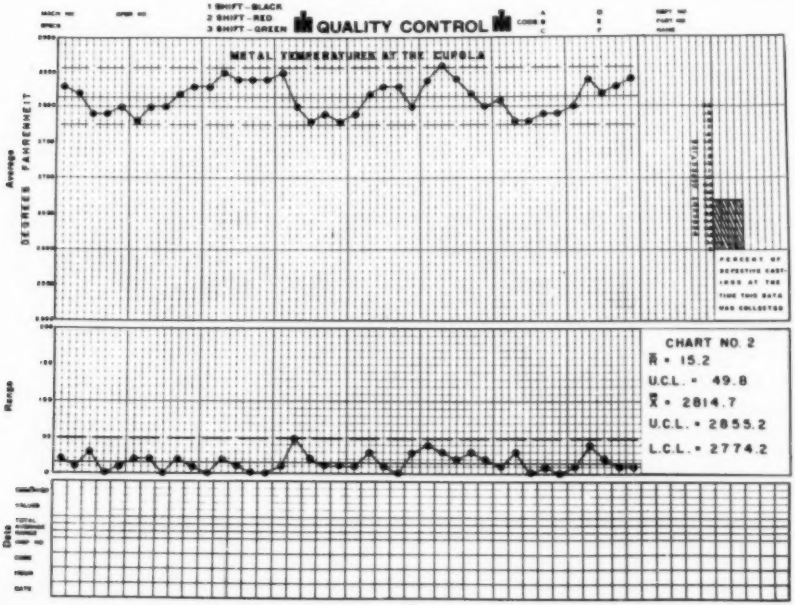

1 SHIFT - BLACK
2 SHIFT - RED
3 SHIFT - GREEN
QUALITY CONTROL
METAL TEMPERATURES AT THE CUPOLA
Average
DEGREES FAHRENHEIT
Range
Data
CHART NO. 2
R̄ = 15.2
U.C.L. = 49.8
X̄ = 2814.7
U.C.L. = 2855.2
L.C.L. = 2774.2

Fig. 6

Molding:

The use of Quality Control either as a process control or acceptance procedure often depends upon a particular foundry layout. A floor molding foundry which only pours off a few hours a day has to rely generally upon a system where castings are checked for defects possibly a day after being poured; and necessary changes in molding practice are made after that time. Control in this case must be based on the overall charting picture of molder performance since the molder puts up his floor almost completely before pouring off the molds; and a check at pouring time is almost useless, for the quality of the unpoured floor is completely unknown. One type of chart that has been of value is an individual molder's P chart (Figure 7).

INDIVIDUAL MOLDER'S P CHART

SCRAP DUE TO MOLDING PRACTICE

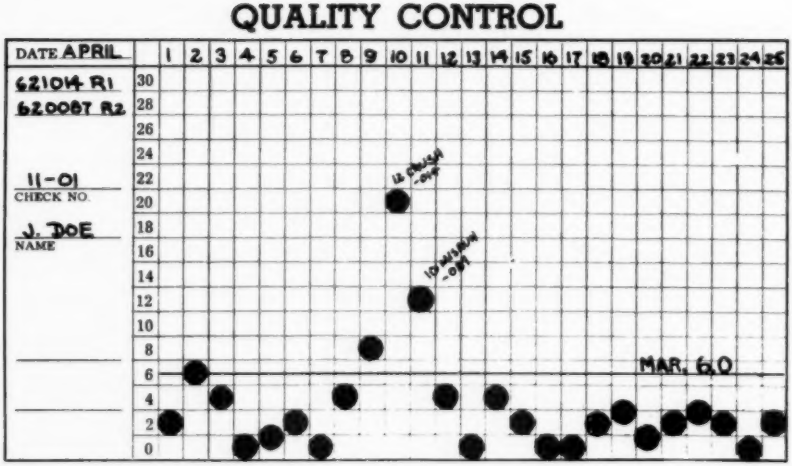


Fig. 7

In one of our smaller foundries we have 70 single and group molders. Each man and group has a separate card mounted on a board near the foundry office. On each card is recorded the per cent of scrap attributable to the molder each day. Being out in plain sight the board shows up the good and bad performances alike, and arouses a sense of competition merely by its presence. Admittedly the charts are not strictly up to date but the continuous story they tell, even when they are two days behind, is a forceful one and cannot be overlooked. Reduction of scrap was noticed shortly after the board had been mounted and it has awakened a desire for better performance in those molders.

In mechanized shops, sample type inspection is being used at the end of rollaway conveyors or at the end of unit shakeouts. Samples can be taken as often as desired and knowledge of specific defects is evident sometimes within 15 minutes after the castings have been poured. The use once again of an individual P chart is very helpful especially if it is posted near the molder's machine, plotted often, and the defects listed.

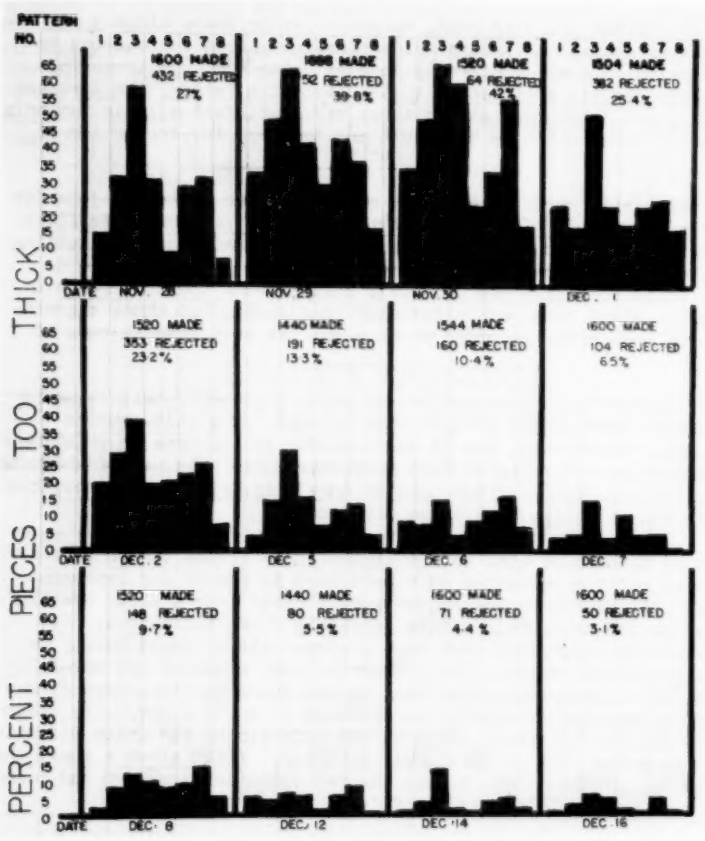
Rating systems are also kept, very similar to those already mentioned on core making operations, where molders are penalized for casting defects and rated on their conformity to a set standard. In these instances where results are plotted in the form of a rating, the issuance of the 1, 2, and 3 colored notices serve very satisfactorily in reminding the molder of scrap which is his fault and the need for correction of same.

Characteristic cards have proved very useful in preparing jobs for the molding floor. By listing the complete specifications as to flask requirements, chills, chaplets, ramming, previous defects, core setting practice, etc., a job can be reviewed each time it is reinstalled to make sure it is correctly set up. This will prevent the occurrence of defects which might have been eliminated previously, but whose corrective measures were never recorded to help out the next time in case of trouble.

Such items as mold hardness can be controlled accurately by means of $\bar{X}R$ charts. Whether the mold is made by hand, on a jolt-squeeze machine, or a sandslinger, the $\bar{X}R$ chart serves its purpose in pointing out the variation occurring on that characteristic. The use of solenoid valves to control "bumping" has been of some help on jolt machines, but alone they cannot entirely cure the problem.

Another example of what can be accomplished by the use of charts was the core setting operation on a seed boot in one of our implement foundries. All machining operations were located from a cored casting surface which had to be held within extremely close tolerances. The molder naturally complained that such a close setting could hardly be kept on a production molding job. Nevertheless, after he was carefully instructed as to the proper core set, he was shown how the cored dimension would be sampled frequently and plotted on an $\bar{X}R$ chart. The chart would show him the tolerance which he was maintaining and which direction he should move the core if it was off location. After about a week, the man was easily setting cores to the desired specification, and had plenty of pride in his own ability and confidence in the chart.

Frequency distributions are another device which can show the overall status of a process at a glance. They are ideal for such items as conformity to analysis specifications and mold hardness distribution where the whole story might be desired rather than a point by point process record. One of our Canadian foundries had molding trouble with a small fluted washer which resulted in about 42% scrap. Quality Control was applied to the job in the form of frequency distributions represented by histograms on the per cent of defective castings produced from each of the washers on the gate. (Figure 8).



X-400 FLUTED WASHER CAST 8.1. CHECK BY SLIDE GAUGE FOR THICKNESS

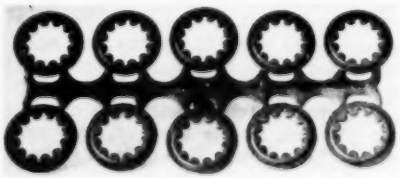


Fig. 8

Samples were taken from the molding floor, and the rejections were segregated under the respective washer numbers. It was found that one washer was accounting for as much as 67% of the total rejections. This one washer, together with others causing scrap to a lesser degree, was gradually corrected with the foreman regulating his practice based on information gained from a daily perusal of the charted scrap results and recorded defects. In this manner, the washer scrap was quickly reduced to a figure of 3.1% in three weeks.

Advanced Techniques:

The use of advanced techniques has been very gradual, for once again we must remember we are dealing with an old art which for years was dependent upon the skill of the individual foundryman. A look at the American Foundrymen's Society book "Analysis of Casting Defects" shows mostly pictures of defects and a large listing of possible causes which are still hotly disputed throughout the industry. We are past the stage where we made investigations by trying to hold all elements constant - varying only one at a time to observe its effect. With some of the advanced methods, the important items can be isolated while observing the variations of all elements during normal operation. Learning to accept these normal variations seems to be most important for they will always exist and we must set the pattern of our operations recognizing that fact.

Correlation is being used frequently by a few of our Quality Control Engineers on specific investigations. Bore hardness of cylinder blocks versus top deck hardness constituted a large program at Indianapolis. The innoculation of iron on the basis of chill test readings was another. Disproving the old axiom of 2½ men per ton of metal poured was a revelation to some of our personnel, for we found we actually were using an "S" shaped relationship.

At our Research foundry in Chicago, advanced studies were made which revised the method of thinking of the whole foundry group. One of the prime results has been the realization of the existence of confidence limits on sets of figures used in the various reports. Many times, comparisons were made and conclusions drawn from too few figures, where the true results might possibly have been far different than they seem. Another investigation involved an analysis of variance problem on the tensile strength of certain sand and resin mixtures for shell molding. By designing the experiment beforehand, results were obtained in a single week, which otherwise might have taken more than a month if normal procedure had been followed.

Admittedly there are not too many places where such investigations are being used, however, their value should not be overlooked. They can often provide definite proof as to the validity of certain practices or theories, they can save time, and they can open new channels for further investigations.

Although most of this material has been concerned with operations in large mechanized foundries, many of the applications can readily be transferred to smaller plants. Quality Control has the appeal of being readily adaptable in some form to almost all setups and can be of definite value if correctly used.

Once it is installed and accepted by both management and the employees, it can be the beginning of a very rewarding experience for all concerned.

BIBLIOGRAPHY

1. "Castings With a Pedigree" - Loper, Claffey, Waldenmeyer - American Foundryman, Feb. 1953 - Vol. 23 No. 2

2. "Melt Department Incentive Plan" - E. G. Tetzlaff - Transactions-American Foundrymens Society, 1952 - Vol. 60 p 432.

3. "Foundry Quality Control" - A. Evans - Industrial Quality Control January 1953 - Vol. IX No. 4

THE AFFECT OF THE CONCENTRATION AND FLAVOR INTENSITY IN A FOOD PRODUCT ON THE ORGANOLEPTIC SENSES

D. A. Brandt, E. S. Edelmann and N. A. Forte
Schenley Distillers, Inc.

The traditional acceptable quality level in the distilling industry in the past was ordinarily determined by an expert, or a small group of experts. Since the need became apparent, in this industry, to use a hammer to drive a nail, scientific methods are gradually replacing the long established subjective procedures. The requirement for an objective control system which would apply to the odor and taste senses was highlighted by the fact that the consumer was becoming increasingly more discriminating and more demanding for the continuity of flavor in his favorite distilled beverage. As modern scientific tools were applied to the control of the uniformity of end product, it became necessary to strengthen and to expand physical and chemical techniques from the beginning of the process; that is, in the purchasing and acceptance of grain until the finished raw product was drawn off the still. Many great advancements were made in the alcohol industry during, and since World War II, both in increasing quantity and in improving quality; yet repeated experiments with physical and chemical tests have failed to produce to date any analysis which compares in reliability with the God given senses of taste and odor.

Even to this day, there is considerable discussion among the initiated in the business in regard to the proper method to determine quality. There is some belief existing at this time, that a test may be conducted either in odor or taste at full bottle proof with reliable results.

Our intentions in preparing this paper were to gather data which would establish with some degree of finality the proper proof at which organoleptic judgment should be made. In designing the experiment, we determined that we would use all of the laboratory facilities established by Schenley Distillers, Inc., for the control of the taste uniformity of products. Four different distilled products were studied. These products were straight whiskey, blended whiskey, brandy and gin. Although 1500 to 1800 individual judgments were obtained, using the duo trio technique for uniformity, we are not prepared to draw any significant conclusion from the data obtained thus far. The results have indicated some variation from laboratory to laboratory, which may be explained from at least two different points of view.

1. All of the laboratories were not trained in advance for the specific items examined. For example, two of the laboratories had had no previous experience with either brandies or gins. Another one of the laboratories was not familiar with blends. As familiarity, obtained by repetition, is a very desirable attribute in organoleptic testing, some consideration must be given to this factor.

2. The usual incidence of respiratory infections in the winter just concluded, has played some role in the lack of consistent ability among even the fully experienced observers.

Each plant laboratory has a panel consisting of approximately 100 members. These range in proficiency from slightly over the 50% correct

point to well into the 70% correct. The yardstick of reliability for a taster is his ability to correctly match the sample with the standard in the duo trio test. This experiment is conducted by presenting a series of seven samples in which quantity, color, proof, temperature, time interval, order of presentation, neutral location and motivation are controlled. The observer is offered first a conditioning sample after which he is told to expectorate the specimen and rinse his mouth. The purpose of the distilled water rinse, of course, is to decelerate the fatigue factor on the taste buds. A 20 second interval elapses from the introduction of the conditioning sample to the reference standard. Ten seconds are allowed between the point of reference and sample one. After a ten second interval, sample two is presented. The taster is then asked to select whether sample one, or sample two, is identical to the standard. The mouth is rinsed with distilled water again and the trio part of the test is repeated after a twenty second pause to regain sensitivity. A running record is maintained of each individual's ability to correctly match the items compared. At weekly intervals he is advised of the total number of observations he has made together with his percentage of correct responses. After a reasonable length of time, unqualified observers are dropped from the panel. The tasters are obtained from office and plant employee groups. Although the average taster does not consider that he has the ability to taste whiskey before he becomes a panelist, he soon convinces himself if he is competent, that he has a keen sense of discrimination.

Two straight whiskies were used for the first products tested in this experiment. A total of 480 judgments was obtained when comparing these at 86 proof. Proof is the measure of alcoholic content in the beverage. Data when analyzed in terms of the standard error of a percentage results in a 3.5 $\sigma_{\bar{p}}$. The presentation of an alcoholic beverage at this proof produces a strong sensation in the mouth. The degree of training at this proof might well affect the discriminating powers of the taste buds. We, however, are familiar with the fact that the greater the quantity of alcohol imbibed, the greater the desensitizing effect on the taste mechanism. The relatively large number of judgments obtained has exerted of course, its affect on the $\sigma_{\bar{p}}$ score. Individual plant laboratories generally had acceptable $\sigma_{\bar{p}}$ ratings at this proof which would indicate a lack of ability to properly identify the whiskies presented at this strength. As a second step, in experiment one, the proof was reduced to 45°, and a total of 580 judgments was obtained. The combined $\sigma_{\bar{p}}$ rating at this proof was 5.6. The various laboratories were able to individually detect differences between the two whiskies at this proof as indicated by higher $\sigma_{\bar{p}}$ ratings at each location. We should like to point out that the proof generally used for comparing distilled products in our laboratories is 45°. It should be emphasized, therefore, that our observers have been trained at this proof and, therefore, are probably better able to detect differences due to past experience alone. To complete experiment one, the proof was further reduced to 25°, and a $\sigma_{\bar{p}}$ rating of 5.5 was obtained on 520 judgments. With one exception, the various laboratories were fairly well able to determine differences between the samples offered. As a general statement, it may be said that straight whiskies contain a relatively greater amount of flavor. There possibly is a threshhold at which a sufficient quantity of flavor is received through taste sensation - yet is not so great that the taste buds are desensitized to such an extent that additional samples are not discernible. We should like to again point out that the data collected on this entire paper are not sufficient to draw any conclusions with certainty; we are unable at this point to definitely establish the optimum proof.

Experiment two was conducted with blended whiskies, and on a total of 430 judgments at 86 proof, combined results showed a 3.3 σ_p. With one exception the laboratories were unable to detect significant differences on an individual basis. This again would indicate that possibly there was too much flavor, or too much alcohol in the whiskies to permit detection of differences. In blended whiskies, generally, flavors are lighter and possibly would exert less desensitizing effect on the taste senses than the alcohol. At 45 proof a total of 440 judgments was obtained. Upon compiling the individual plant results a σ_p score of 4.6 resulted. The third portion of experiment two was conducted at 25 proof. The overall total of plant judgments was 430 with a resultant 3.0 σ_p. The trend, therefore, on both experiment one and experiment two, that is with the straight whiskey and the blended whiskey, would pose an objection to the use of 86 proof for organoleptic testing. The 45 proof, in both instances, indicated the highest σ_p rating, while on the straight whiskies almost an identical σ_p rating was obtained at 25 proof. We should like to further state, that, although differences were present between both the two straight whiskies utilized and the two blends, the measure of difference on a relative scale was small. Lack of familiarity with the products tested was also a factor to be considered. The plant with the least experience in blended whiskies, for example, reported the lowest σ_p score at all three proofs examined.

Experiment three was conducted with two samples of brandy. On 570 judgments obtained at 84 proof, a 4.6 σ_p was calculated from all sources. Three of the plants showed relatively good differentiation at this proof. Two of these three plants are regularly concerned with process control on brandy: the other three locations were not as well able to detect differences on these two samples. One location which had had no previous experience with brandy, consistently reported the lowest σ_p ratings. At 45 proof a σ_p of 4.6 was obtained on the basis of 580 individual observations. At 25 proof 640 observations resulted in a 5.1 σ_p. Lack of any definite pattern on this experiment makes it difficult to draw any conclusion from this data. If experience with the product tested were to be considered, and the judgments eliminated from the plant with no previous experience, a higher σ_p rating would be obtained for the 25 proof, and, therefore, some consideration should be given to the use of this proof - at least for experimental purposes.

The fourth product tested on a uniformity basis was gin. A total of 540 observations were made at 90 proof. When calculating the σ_p score, the figure 4.4 evolved. At 45 proof 5.5 σ_p resulted from the analysis of 560 judgments. Upon reducing the gins to 25 proof, a 5.7 σ_p resulted from 570 judgments. The greatest individual plant discriminating ability on gins occurred at 25 proof. The first four experiments were all conducted at the various proofs with the duo-trio technique.

In order to vary our techniques we used the triangle method for obtaining our results on the odor tests. The procedure utilized was the presentation of three different 250 milliliter red tinted Erlenmeyer glass stoppered flasks which contained 50 milliliter of sample. The observer was instructed that he was to select the two out of the three samples which were the same on an odor basis. He was permitted to return to each sample as often as he desired. The same four distilled products were compared on odor as they had previously been checked on taste.

The fifth test was conducted on straight whiskies at 86 proof. A

total of 600 judgments resulted in 3.1 σ_p score. A 2.9 σ_p was obtained on 600 judgments at 45 proof, while a 2.7 σ_p was the result of 600 judgments at 25 proof. Although no selection can be made from the overall results on the odor tests of straight whiskies at this point, greatest consistency was registered among the individual laboratories at 45 proof.

The blended whiskies when compared at 86 proof on 500 judgments indicated a 1.7 σ_p. The combined results from all plants at 45 proof, again on 500 individual observations show a 2.9 σ_p. At 25 proof a 3.4 σ_p was the result. The great variation exhibited on the blend odor studies does not permit us to select any proof as showing a trend. The 2.9 σ_p and 3.4 σ_p, however, which resulted from the tests at 45° and 25° indicate that these proofs possibly warrant additional study. We should like to state at this point that odor tests at Schenley are utilized only as auxiliary control procedures, and are not a part of regular control work. As an additional explanation for inconsistency, we offer the fact that the triangle technique was used at almost all locations for the first time on this project. It is our intention to continue to work with the particular method on both odor and taste to see whether or not one method can be shown to be more suitable for our purpose than the other.

The brandies were compared at 84 proof, 45 proof and 25 proof, with ratings of 6.0 σ_p, 3.6 σ_p and 4.7 σ_p respectively.

The gin was examined at 90 proof, 45 proof and 25 proof with the resultant ratings of 4.8 σ_p, 7.0 σ_p and 7.4 σ_p. The gin as a group showed greater consistency in odor than any group tested. If we are allowed to discard the results of one of the laboratories, it would appear that good discriminating ability was evident at all three proofs. Brandies and gins generally are more aromatic in odor. It is possible that smaller differences on these two products may be detected on an odor basis.

The results of both the taste and odor tests indicated that when considering the data collected from all sources, the samples of each of the four products tested differed from each other at all proofs. As we previously indicated the scores obtained at various proofs at individual locations did not fit in with the overall pattern established. The indications are that the suggested proofs mentioned-especially on the taste tests-may be profitably explored to a greater extent. Our present plans call for a restudy on the same products during coming months when weather conditions will be more conducive to physical capabilities. The additional studies should also be of greater significance due to the fact that the initial experience has been gained through the performance of these experiments. We shall report, therefore, at a future date the results of what we hope to be conclusive studies on this important subject; and through which we may offer suggestions as to not only the proper flavor intensity and proof, but also to the relative merits of the duo-trio and triangle technique.

ACKNOWLEDGEMENTS

Recognition is due James F. Banks, Geo. T. Stagg Company, Frankfort, Kentucky; David Dewar, Bernheim Distilling Company, Louisville, Kentucky; Paul Hutchison, Schenley Distillers, Inc., Lawrenceburg, Indiana; Joseph C. Mueller, Schenley Distillers, Inc., Fresno, California and Richard Ousler, Joseph S. Finch and Company, Schenley, Pennsylvania, for their aid and perseverance in collecting data for this project.

BIBLIOGRAPHY

A. Statistical Methods for Quality Control.

(1) A First Course in Statistics, E. F. Lindquist. Riverside Press, Cambridge, Massachusetts, 1938.

(2) Techniques of Statistical Analysis, Statistical Research Group. Columbia University, McGraw-Hill Book Co., New York, 1947.

(3) Statistical Quality Control, E. L. Grant. McGraw-Hill Book Co., New York, 1946.

(4) Experimental Designs, W. G. Cochran and G. M. Cox, Wiley, 1950.

(5) Tests of Significance, What They Mean and How to Use Them, J. H. Smith, University of Chicago Press, Chicago.

B. Psychometrics

(1) Properties of Foods. E. C. Ziegler, and E. H. Scofield, Food Manufacture, 20, 1945, 55.

(2) Preferential Odor Rating of the Essential Oils of Coriander Seeds. E. H. Scofield, National Farm Chemurgic Council Bulletin, August, 1940.

(3) Some Approaches to the Measurement of Taste and Related Properties of Food and Beverages. E. H. Scofield, Conference Course Lectures, School of Business, University of Chicago, 1948, p 25-31.

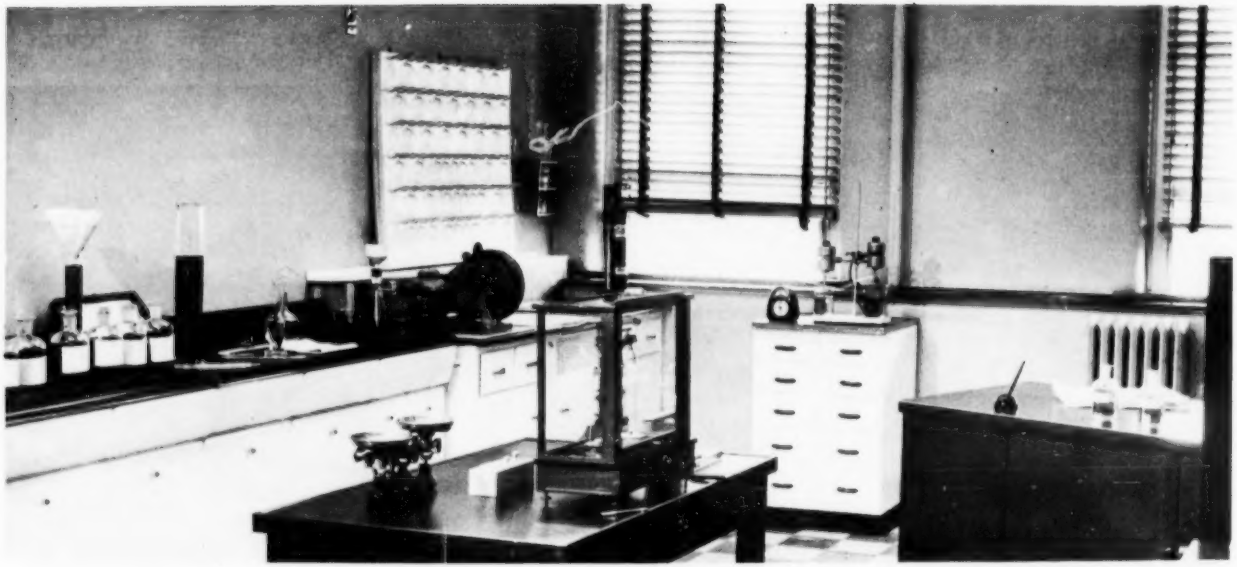
(4) Measurements of Sensory Differences, David R. Peryam, and VeNona W. Swartz, Food Technology, IV, No. 10, 1950, 390-395.

(5) Sensory Methods for Measuring Differences in Food Quality, Elsie H. Dawson and Betsy L. Harris, Agriculture Information Bulletin No. 34, United States Department of Agriculture, Washington, D. C. 1951.

C. Current Literature

(1) Industrial Quality Control, American Society for Quality Control, Alfred L. Davis, Executive Secretary, Room 5036, 70 East 45th Street, New York 17.

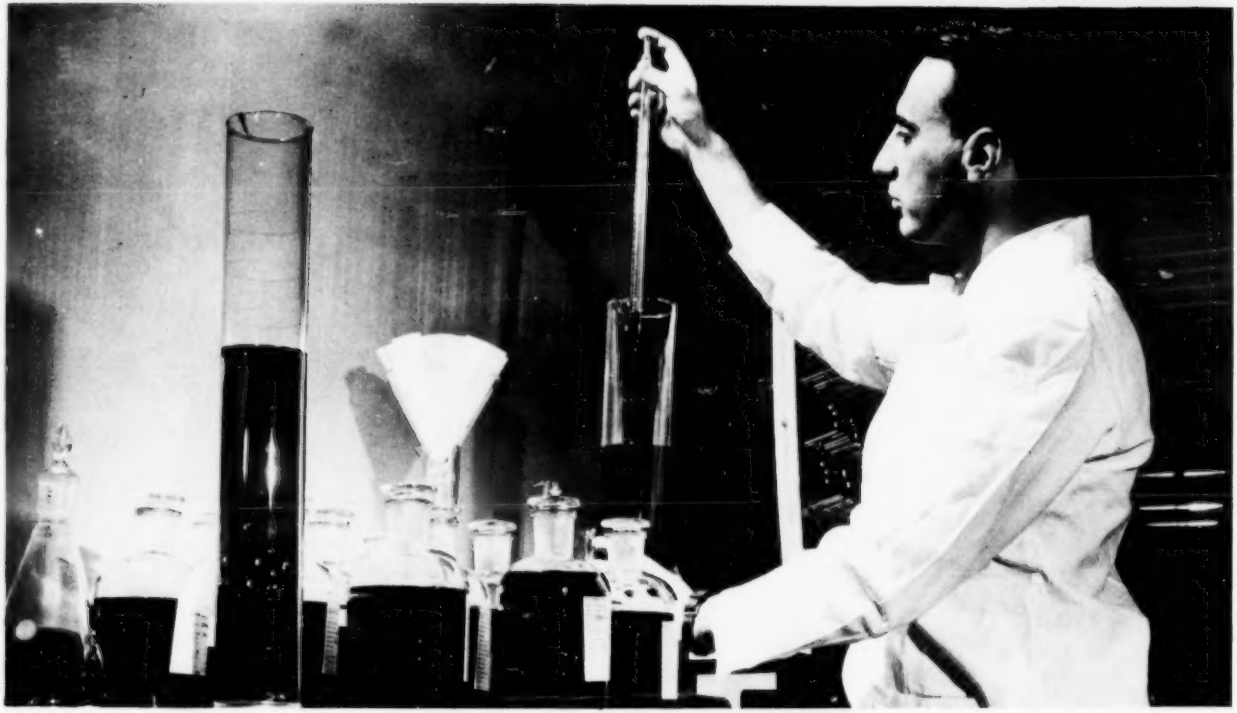
Formulas (such as pictured center) for our experimental blends of whiskies are based on the maturing inventories at Schenley's numerous distillery warehouses. The wide selection of various whiskey types is illustrated (on the left) by the summary of the bulk inventory.



General view of Schenley's experimental blending laboratory at its Cincinnati Central Control headquarters. The most modern laboratory equipment has been provided so that ideal conditions exist for the proper preparation of all experimental whiskey blends. The apparatus is so extensive that even a small water deionizer is utilized to afford the lab with a ready source of fresh, pure water for reducing proof. The bottles seen at left are made of pyrex and fitted with ground glass stoppers to eliminateany possible affect of the glass on the whiskey after the latter has been withdrawn from the barrel. Through the utilization of such "control" properties, Schenley is assured that at all times it has samples in its whiskey library which are truly representative of its aging stocks.

A great variety of whiskies is maintained on file in specially-produced glass bottles at Schenley's Cincinnati Central Control Laboratory. These samples, used for experimental blending purposes, are replenished at regular intervals by the withdrawal of aging whiskies from barrels stored in the company's distillery warehouses.

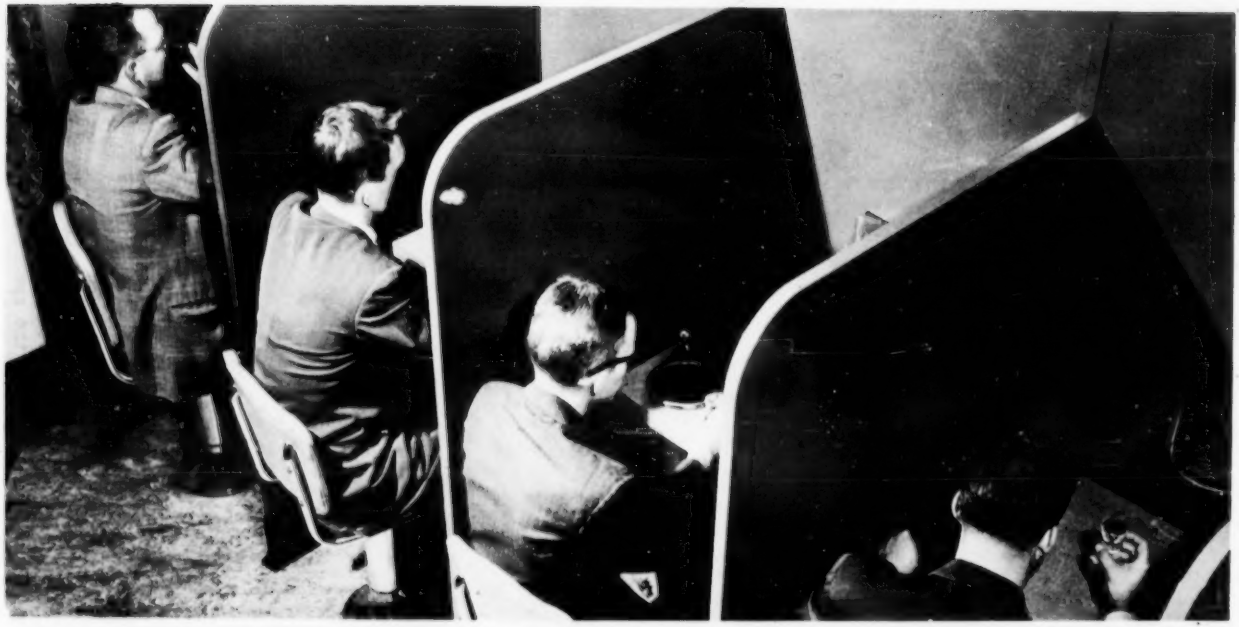
A lab technician is shown withdrawing one of the whiskies to be used in the preparation of an experimental blend. He is referring to a blending formula which is based on matured stocks of whiskies available for bottling.



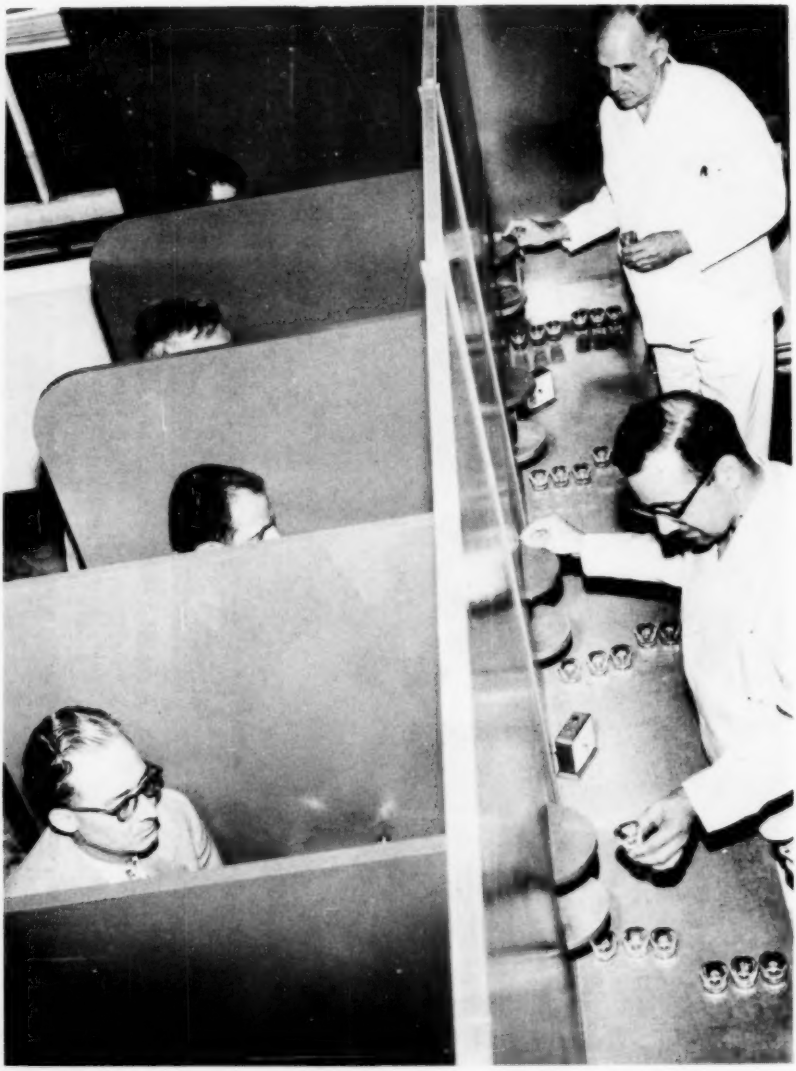
A lab technician carefully measures out various whiskies obtained from our extensive library in the Cincinnati Central Control Laboratory. In this operation, he is preparing experimental blends for later comparison with a given standard. Note that he is using calibrated equipment to accurately ascertain exact quantities of each component.

Close-up of a lab technician drawing an exact quantity of whiskey into a shot glass in preparation for a taste uniformity test. It is evident from the tray pictured that random order presentation of samples is utilized in this technique. The box is numbered and it can be noted that the individual shot glasses are lettered "S" and "U". In addition, each group of seven glasses has a number which corresponds to the individual booth in which that group of samples will be used. Each taster therefore receives his samples in different order so that the test is truly objective, i.e., panel members seated in adjoining booths taste the same two whiskies but do not receive them in the same order.

The tasting booth in the Central Control Laboratory as viewed from the rear. Here, two of the lab assistants are at work conducting a taste uniformity test. Operating on a time-controlled schedule (note the special clock), they submit the whiskey samples to the panel members seated at the front of the booth. As the lab men deliver filled samples through the turn-table mechanisms seen in the photo, the tasters in the same movement return the emptied glasses. This assures a continuity necessary for compliance with the established time "control" conditions.



Still another view of the tasting booth in Schenley's Cincinnati Central Control Laboratory. The tasters never actually consume the whiskey samples. They merely retain the whiskey in their mouths long enough to carefully judge its taste, then expectorate it into the wells, such as can be seen in the second booth from the right. Each booth also is equipped with distilled water so that the panel members can rinse their mouths when so instructed. These "rinses" serve to freshen each member's taste buds for the opening or follow-up series of tests. Two such tests are given at each sitting.



Another view of the tasting booth in Schenley's Cincinnati Central Control Laboratory. Members of a tasting panel seated at left are served the whiskey samples to be tested through turn-table openings such as shown in front of the lab technicians at right. From their side of the booth, the technicians, operating unseen and on a split-second time schedule, submit the samples to the panelists. Time of the tests is controlled through use of the specially-equipped clocks pictured, so as to assure maximum sensitivity of the panel member's taste buds.

A view of Schenley's Central Quality Control Laboratory in Cincinnati. In the background, members of an employee tasting panel are taste-testing samples of whiskey being submitted to them by Lab technicians who are on the other side of the booth.

THE APPLICATION OF AN IBM 604 ELECTRONIC CALCULATING PUNCH TO SOLVE MULTIPLE LINEAR REGRESSION PROBLEMS

D. H. W. Allan and R. F. Attridge
The Steel Company of Canada, Limited

When we have a mass of data involving more than two variables, a sound way of assessing the effect of the independent variables on the dependent variable is by the technique of multiple linear regression. As we all know, this technique entails considerable computational labour and invariably requires a lot of time. At Stelco there existed only a limited amount of time to devote to multiple linear regression problems, and so it was decided to investigate the possibility of placing the bulk of the computational labour on the IBM punch-card equipment in the Tabulating Department. It is the purpose of this paper to outline a method of adapting the solution of multiple linear regression problems to a punch-card procedure designed to make maximum use of the IBM 604 Electronic Calculating Punch.

It should be remarked that the development of any punch-card procedure is limited not only to the type of equipment in the Tabulating Department, but also to those machines on which spare time is available. This statement is made since to date most tabulating installations are being used mainly for accounting purposes, and as a result, any technical computations must fit into the 'off peak' periods of the accounting schedule. At Stelco the IBM 604 was the machine most available and, accordingly, the procedure developed made as much use as possible of this unit.

Since no members of our Metallurgical Department had ever had previous experience with punch-card methods, it was necessary to investigate how other people were employing such methods to solve similar problems using IBM punch-card equipment. Therefore, a task force, consisting of our Tabulating Department Supervisor, the local IBM office manager, an IBM salesman specializing in computation, and the writers, was formed to develop an acceptable punch-card procedure.

The first decision of the task force was the adopting of the computational procedure for solving multiple linear regression problems developed by Paul G. Hoel in Chapter VII of his excellent book "Introduction to Mathematical Statistics."(1) The major reasons for adopting Hoel's procedure was that it involved the computing of quantities, whose computation could be easily adapted to punch-card methods and that once these quantities are computed the least squares regression equation, the multiple correlation coefficient, the partial correlation coefficients and the standard error of estimate are readily calculated on a desk calculator.

Outline of Hoel's Procedure

To illustrate Hoel's procedure let X_0 denote the dependent variable and $X_1, X_2, \cdots, X_k$ the k independent variables, and suppose that a sample of size n is available for these $k+1$ variables. It is assumed we wish to estimate the dependent variable X_0 by means of a linear function of the independent variables. Thus, if the estimated value of X_0 is denoted by X'_0, the relationship may be expressed as

(1) $$X_o' = c + c_1 X_1 + c_2 X_2 + \cdots + c_k X_k \ .$$

The constant c and the coefficients c_i, $i = 1, 2, \cdots, k$, are of course unknown and must be determined by means of the available data. These unknown c's are determined by the method of least squares; that is, the c's determined are those that minimize the sum of squares

$$\sum (X_o - X_o')^2 \ ,$$

where the sum extends over the n sample values.

Equation (1) is often referred to as the least squares regression equation or more simply as the regression equation.

In order to simplify the solution of the problem Hoel expresses all quantities obtained from the data in terms of the familiar statistic, the simple correlation coefficient. The simple correlation coefficient between the variables X_i and X_j is denoted by r_{ij} and is defined by

(2) $$r_{ij} = \frac{\sum (X_i - \bar{X}_i)(X_j - \bar{X}_j)}{n \sigma_i \sigma_j} \ ,$$

where σ_i and σ_j represent the standard deviations of variables i and j respectively and in particular

(3) $$\sigma_i = \sqrt{\frac{\sum (X_i - \bar{X}_i)^2}{n}} \ .$$

It should be noted that $r_{ij} = r_{ji}$ and also if $j = i$, that $r_{ij} = r_{ii} = 1$.

After these simple correlation coefficients are computed for all different pairs of variables, they are arranged in the following square array called a matrix:

(4) $$\begin{bmatrix} r_{oo} & r_{o1} & r_{o2} & \cdots & r_{ok} \\ r_{1o} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{2o} & r_{21} & r_{22} & \cdots & r_{2k} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{ko} & r_{k1} & r_{k2} & \cdots & r_{kk} \end{bmatrix} \ .$$

Since by (2) $r_{ij} = r_{ji}$, then (4) is further called a symmetric matrix.

The next step in the procedure is to compute the determinant of the matrix (4), namely

(5) $$R = \begin{vmatrix} r_{oo} & r_{o1} & r_{o2} & \cdots & r_{ok} \\ r_{1o} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{2o} & r_{21} & r_{22} & \cdots & r_{2k} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{ko} & r_{k1} & r_{k2} & \cdots & r_{kk} \end{vmatrix} \ ,$$

and what are called the cofactors of each element r_{ij} in the determinant. Before defining the cofactor of r_{ij} we must first define the minor of r_{ij}. The minor of an element r_{ij} is defined as the determinant obtained by stroking out row i and column j of the determinant R. The cofactor of an element r_{ij} is defined as $(-1)^{i+j}$ times the minor of r_{ij} and is usually denoted by the corresponding capital letter R_{ij}.

Using the method of least squares Hoel shows that the least squares regression equation becomes

$$X'_o = \sum_{i=0}^{k} \frac{\sigma_o}{R_{oo}} \frac{R_{oi}}{\sigma_i} \bar{X}_i - \frac{\sigma_o}{R_{oo}} \frac{R_{o1}}{\sigma_1} X_1 \tag{6}$$

$$- \frac{\sigma_o}{R_{oo}} \frac{R_{o2}}{\sigma_2} X_2 - \cdots - \frac{\sigma_o}{R_{oo}} \frac{R_{ok}}{\sigma_k} X_k .$$

Comparing this equation with (1) we see that

$$c_i = - \frac{\sigma_o}{R_{oo}} \frac{R_{oi}}{\sigma_i} , \qquad i = 1, 2, \cdots, k, \tag{7}$$

and

$$c = - \sum_{i=0}^{k} c_i \bar{X}_i . \tag{8}$$

Hoel next demonstrates that the standard error of estimate, which we shall denote by $\sigma_{(x_o - x'_o)}$ is given by the formula

$$\sigma_{(x_o - x'_o)} = \sigma_o \sqrt{\frac{R}{R_{11}}} . \tag{9}$$

The multiple correlation coefficient, which is denoted by $r_{o \cdot 12 \cdots k}$ to distinguish it from the simple correlation coefficient, serves as a measure of the usefulness of the regression equation for estimating purposes. This coefficient is given by the formula

$$r_{o \cdot 12 \cdots k} = \sqrt{1 - \frac{R}{R_{11}}} . \tag{10}$$

Finally, Hoel obtains a formula which serves as a definition of the partial correlation coefficient. The partial correlation coefficient between two variables is the correlation coefficient between the two variables when the remaining variables are held fixed. Unlike Hoel, we shall denote the partial correlation coefficient between variables i and j by p_{ij}. The formula obtained by Hoel is

$$p_{ij} = - \frac{R_{ij}}{\sqrt{R_{ii} R_{jj}}} . \tag{11}$$

Modification of Formulae for Punch-Card Procedure

As the data obtained in industry is rarely accurate to more than four digits, it was decided to limit all numbers to four significant digits. Hence any numbers accurate to more than four digits are rounded off to four significant digits.

The handling of the original data on the punch-cards is simplified if the magnitude of the numbers for each variable is made as near the same order as possible. To do this, the original data for each variable is multiplied by a power of ten such that all numbers fit into a four position field with the left hand significant digit of the numerically largest number in the high order position of the field and the decimal point at the right of the low order position of the field. The power of ten is called the scale factor and denoted by v_i. Thus if we denote the original value of the variable by Y_i, and the adjusted value of the variable by X_i, then

$$Y_i v_i = X_i , \qquad i = o,1,\cdots,k. \tag{12}$$

The computation is carried out using these adjusted values of the variables and then at the finish a readjustment is easily made to restore the decimal points to their original position.

In order to be able to use the simple forms of the formulae for the standard deviation and the simple correlation coefficient, it was decided to transform the data for each variable into deviations from the mean. It was further decided to multiply all the deviations by the reciprocal of the numerically largest deviation so that the numerically largest deviation becomes plus one or minus one, depending on its sign, and all other deviations lie within the range plus one to minus one. Only a relatively few of the multiplied deviations will have a numerical value greater than 0.99999, i.e. 1.00000; hence the high order position of the field allowed for the multiplied deviations would contain zero in most cases. Since this high order position would carry little essential information and carrying it would reduce by one the number of digits in the result, the 604 calculate control panel for the calculation of the multiplied deviations has been wired to test the size of the result obtained in each case and if this result is 1.00000 or -1.0000 to replace it respectively by .99999 or -.99999. Whereas before six positions would be required to carry the information, now only five positions are required and we are able to make maximum use of the storage capacity in the 604. The above operation introduces a small error, but it increases the overall accuracy for these calculations as one extra place to the right of the decimal point can be carried.

This process of transforming the original data into deviations lying within the range plus one to minus one is called coding the variables. The reciprocal of the numerically largest deviation is called the coding factor and denoted by f_i. Denoting the multiplied deviations by y_i, $i=o,1,\cdots,k$, then we have

$$(X_i - \bar{X}_i) f_i = y_i , \qquad i = o,1,\cdots,k. \tag{13}$$

Thus in terms of the coded variable the formula for the standard deviation is

(14) $$\sigma_i = \sqrt{\frac{\sum y_i^2}{f_i^2\, n}} = \frac{1}{f_i}\sqrt{\frac{\sum y_i^2}{n}} \quad .$$

Using equations (13) and (14) we can easily show that the simple correlation coefficient

(15) $$r_{ij} = \frac{\sum y_i\, y_j}{\sqrt{\sum y_i^2}\ \sqrt{\sum y_j^2}} \quad .$$

This is the form of the equation for the simple correlation coefficient actually used in the calculations on the 604.

Instead of computing the cofactors of each of the elements in the determinant (5) directly, we compute what is called the inverse matrix of the matrix (4). Since the matrix (4) is symmetric, then the inverse of (4) is the matrix whose elements consist of the cofactors of the elements r_{ij} divided by the value of the determinant R. Thus, if R^{ij} denotes the element in row i and column j of the inverse matrix, then

(16) $$R^{ij} = R_{ij} \,/\, R \quad .$$

Using relations (14) and (16), we can easily show that the formulae for the unknown c_i's, $i = 1, 2, \cdots, k$ in the least squares regression equation can be written as

(17) $$c_i = -\frac{\sqrt{\sum y_o^2}}{f_o\, R^{oo}}\ \frac{f_i\, R^{oi}}{\sqrt{\sum y_i^2}} \quad , \qquad i = 1, 2, \cdots, k.$$

Similarly the formula for the standard error of estimate is

(18) $$\sigma_{(x_o - x'_o)} = \sqrt{\frac{\sum y_o^2}{f_o^2\, n\, R^{oo}}} \quad .$$

The least squares regression equation as it now stands is expressed in terms of the adjusted variables. If it is expressed in terms of the original variables it reduces to

(19) $$Y'_o = -\frac{1}{V_o}\sum_{i=0}^{k} c_i \bar{X}_i + \frac{c_1 V_1}{V_o}\, Y_1 + \frac{c_2 V_2}{V_o}\, Y_2 + \cdots + \frac{c_k V_k}{V_o}\, Y_k \quad .$$

The standard error of estimate of the original variable Y_o is

(20) $$\sigma_{(Y_o - Y'_o)} = \frac{1}{V_o}\sqrt{\frac{\sum y_o^2}{f_o^2\, n\, R^{oo}}} \quad .$$

In terms of the elements of the inverse matrix the formulae for the multiple and partial correlation coefficients are respectively

$$r_{0\cdot 12\cdots k} = \sqrt{1 - \frac{1}{R^{00}}}\ , \tag{21}$$

$$p_{ij} = -\frac{R^{ij}}{\sqrt{R^{ii}\,R^{jj}}}\ . \tag{22}$$

Outline of Punch-Card Procedure

With the procedure as it now stands we are restricted to multiple linear regression problems involving not more than ten variables, in other words to problems with not more than nine independent variables. The maximum number of variables could be extended but it was felt that the more complicated control panel circuits and elaborate switches necessary plus the extra expense involved was not warranted. Besides this restriction on the number of variables involved in a problem, there is also a restriction, although it is not as definite, on the number of tests. This restriction on the number of tests arises from the possibility of the sums of squares of the coded variables overspreading the capacity of the storage units in the 604. If the number of tests is less than 1000 there is absolutely no danger of overspreading the storage capacity. Problems with as many as 3000 tests have been handled successfully since the numerical values of the coded variables, which range from 0 to .99999, are rarely clustered about the high end of the range, but are usually evenly distributed .

Although it was decided to use the 604 wherever possible, it was found more efficient for operations consisting of addition only to use the IBM 402 Accounting Machine as it has a higher card feed rate and more counters.

The punch-card procedure as outlined here makes use of the key punch machine, verifier, card sorter, collator and reproducer besides the 402 and the 604 (which includes the 521 gang-punch).

After the data for a problem is collected the first step in the procedure is to adjust the data so the numbers fit into a four position field as explained earlier. A list of the adjusted values of the variables for each test is then submitted to the tabulating department for processing. The main steps in the punch-card procedure are as follows:

(1) The adjusted values of the variables are key punched on raw data cards, one card for each set of test values.

(2) The raw data cards are verified.

(3) The raw data cards are consecutively numbered and gang-punched with the problem number on one run through the 604.

(4) The maximum and minimum values of each variable are obtained using the card sorter. This information is required for the calculation of the coding factors.

(5) The sums of the values of the adjusted variables are obtained on the 402.

(6) Coded values of the variables are computed and punched on new cards called the coded variable cards on successive runs through the 604. One run is required for each variable.

(7) The sums of the cross products of the coded variables are computed on successive runs through the 604. One run is required for every two sums of cross products to be computed.

(8) The square roots of the sums of squares obtained in step (7) are computed on the 604.

(9) The simple correlation coefficients are then computed on the 604 for all pairs of variables. These coefficients are the elements of the matrix of which the inverse matrix is required.

(10) The matrix is inverted using a procedure patterned after one described by John Lowe of Douglas Aircraft at the IBM Computation Seminar in December 1949 (2). This procedure, which is an adaption of the basic elimination method for inverting matrices is carried out in reduction cycles, there being a cycle for each variable, and uses the 604, a card sorter and a reproducer.

(11) Finally, the following reports are printed from the cards by the 402:

(1) Coding Factors - f_i .

(2) Sums of Adjusted Variables - $\sum x_i$.

(3) Square Roots of the Sums of Squares of Coded Variables - $\sqrt{\sum y_i^2}$.

(4) Elements in the Inverse Matrix - R^{ij} .

These reports contain the information required for calculating the least squares regression equation, the multiple and partial correlation coefficients and the standard error of estimate using a desk calculator.

Solution of a Typical Problem

Let us consider the application of this procedure to a typical metallurgical problem.

The problem used here as an example, was undertaken to attempt to answer the following questions: (1) What is the effect of ladle chemistry and thickness on the ultimate tensile strength of plates rolled on the plate mill and (2) what percentage of the variation in the plate ultimate tensile strength is explained by the ladle chemistry and plate thickness. The data for this problem was obtained from 3056 plate tests covering the full range of chemistry and thicknesses of plates produced during a given period.

The adjusting of the data for key punching on the cards is shown in Table I.

TABLE I

Adjusting of Data - Problem 5210

Variable Name	No.	Column Nos.	Test Value	Scale Factor	Field Punched
Ultimate Tensile (p.s.i.)	0	5-8	69940	10^{-1}	6994
% Carbon	1	9-12	0.240	10^{4}	2400
% Phosphorus	2	13-16	0.018	10^{5}	1800
% Sulphur	3	17-20	0.039	10^{5}	3900
% Manganese	4	21-24	0.460	10^{3}	0460
% Silicon	5	25-28	0.073	10^{4}	0730
Thickness (inches)	6	29-32	0.274	10^{3}	0274

As an illustration of the manner in which the raw data cards are key punched, the data of Table 2 is shown key punched on a raw data card in Figure I. The number in columns 1-4 of the card gives the problem number and the number in columns 77-80 the number of the test.

On the completion of the punch-card procedure the following reports were received from the tabulating department.

1. Coding Factors - Problem 5210

i	f_i
0	.000308588
1	.000588831
2	.000580834
3	.000359436
4	.001357936
5	.000549350
6	.000729470

2. Sums of Adjusted Variables - Problem 5210

i	$\sum x_i$
0	19037145
1	05444938
2	05434600
3	10750600
4	01477846
5	03024427
6	01830977

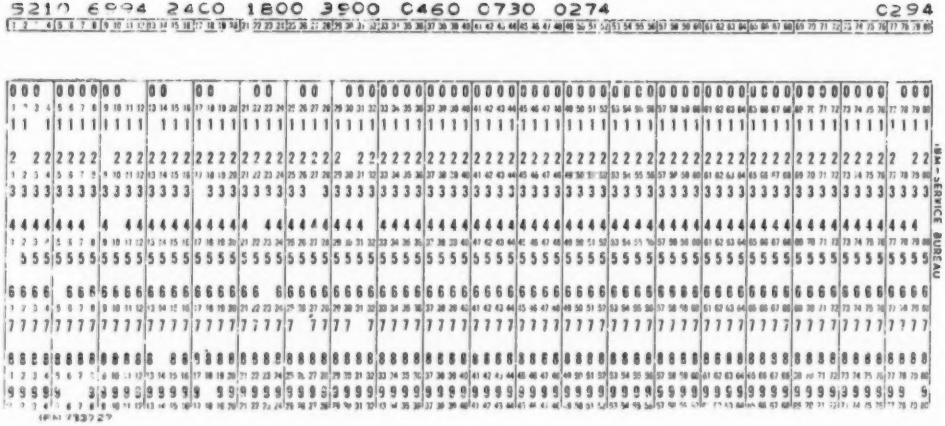


FIGURE I

Key Punched Raw Data Card

3. Square Roots of the Sums of Squares of Coded Variables - Problem 5210

i	$\sqrt{\sum y_i^2}$
0	10.227108
1	14.702064
2	4.846582
3	8.152822
4	9.529968
5	20.220881
6	15.548242

4. Elements of the Inverse Matrix - Problem 5210

Row i	Column j	R^{ij}	Row i	Column j	R^{ij}
0	0	5.0465390	4	0	-1.6175554
0	1	-2.9492035	4	1	0.0084201
0	2	0.0010543	4	2	-0.1438368
0	3	-0.2713333	4	3	-0.0459847
0	4	-1.6175564	4	4	2.6483909
0	5	-1.1826647	4	5	-0.4682166
0	6	1.2230191	4	6	-0.4520824
1	0	-2.9492025	5	0	-1.1826652
1	1	3.4336958	5	1	0.7409952
1	2	-0.0931506	5	2	-0.1466133
1	3	0.0535288	5	3	0.2362264
1	4	0.0084202	5	4	-0.4682166
1	5	0.7049943	5	5	1.8461118
1	6	-1.0529191	5	6	-0.4937388
2	0	0.0010550	6	0	1.2230184
2	1	-0.0931508	6	1	-1.0529192
2	2	1.1154509	6	2	0.0373167
2	3	-0.1683841	6	3	0.0155596
2	4	-0.1438366	6	4	-0.4520829
2	5	-0.1466139	6	5	-0.4937379
2	6	0.0373170	6	6	1.4979022
3	0	-0.2713334			
3	1	0.0535289			
3	2	-0.1683842			
3	3	1.0818823			
3	4	-0.0459847			
3	5	0.2362267			
3	6	0.0155593			

The calculation of the c's in the least squares regression equation is illustrated in Table 2.

TABLE 2

Calculation of the Least Squares Regression Equation - Problem 5210

i	Row 0 of Inverse Matrix R^{oi}	Coding Factor f_i	$\sqrt{\sum y_i^2}$	$\frac{f_i R^{oi}}{\sqrt{\sum y_i^2}}$	$-c_i = \frac{f_i R^{oi}}{\sqrt{\sum y_i^2}} \frac{\sqrt{\sum y_o^2}}{f_o R^{oo}}$
				$\times 10^{-4}$	
0	5.0465390	.000308588	10.227108	1.522719	1
1	-2.9492035	.000588831	14.702064	-1.181183	- .7757065
2	0.0010543	.000580834	4.846582	.0012635	.00082977
3	-0.2713333	.000359436	8.152822	- .1196236	- .07855921
4	-1.6175564	.001357936	9.529968	-2.304875	-1.513657
5	-1.1826647	.000549350	20.220881	- .3213000	- .2110041
6	1.2230191	.000729470	15.548242	.5737985	.3768249

i	$\sum x_i$	$\bar{x}_i$	$-c_i \bar{x}_i$	Scale Factor v_i	$c_i v_i$	$\frac{c_i v_i}{v_o}$
0	19037145	6229.432	6229.432	10^{-1}	-	-
1	05444938	1781.721	-1382.093	10^4	7757	77570
2	05434600	1778.338	1.476	10^5	-89.98	- 829.8
3	10750600	3517.866	- 276.361	10^5	7856	78560
4	01477846	483.588	- 731.986	10^3	1514	15140
5	03024427	989.669	- 208.824	10^4	2110	21100
6	01830977	599.142	225.772	10^3	-376.8	- 3768
	$-\sum_{i=o}^{k} c_i \bar{x}_i = 3857.416$					

If the appropriate values from this table are inserted in expression (19) the least squares regression equation becomes

$$Y_o' = 38600 + 77600\, Y_1 - 830\, Y_2 + 78600\, Y_3 + 15100\, Y_4 + 21100\, Y_5 - 3770\, Y_6 . \tag{23}$$

Substituing in formula (20) we have that the standard error of estimate

$$\sigma_{(Y_o - Y_o')} = \frac{1}{10^{-4}} \sqrt{\frac{104.59373}{(.000308588)^2 \cdot (3055) \cdot (5.046539)}} = 2570 .$$

From forumla (21) we get that the value of the multiple correlation coefficient

$$r_{o \cdot 123456} = \sqrt{1 - \frac{1}{5.046539}} = 0.885 .$$

Finally, we compute the partial correlation coefficients. For example, substituting in formula (22) we have that the partial correlation coefficient between variable 0 and 1 is

$$P_{01} = - \frac{2.949204}{\sqrt{(5.046539) \cdot (3.433696)}} = 0.708 .$$

All the partial correlation coefficients computed are listed in Table 3.

TABLE 3

Partial Correlation Coefficients - Problem 5210

	Ultimate 0	C 1	P 2	S 3	Mn 4	Si 5	Thickness 6
0	1	.708	-.0004	.116	.442	.387	-.445
1		1	.048	-.028	-.003	-.294	.464
2			1	.153	.084	.102	-.029
3				1	.027	-.167	-.012
4					1	.212	.227
5						1	.297
6							1

The equation (23) shows the effect of ladle chemistry and thickness on the ultimate tensile strength of rolled plates. With this equation and given the ladle chemistry and thickness of the plate, we can estimate the ultimate tensile strength of the plate accurate to within plus or minus 5300 ps.i. 95% of the time. It is unwise, however, to use this equation for estimating for values of the independent variables outside the range of the values occurring in the data used to calculate the equation.

The square of the multiple correlation coefficient times 100 may be interpreted as the percentage of the variation in the dependent variable accounted for by the least squares regression equation. In this problem 80% of the variation in the ultimate tensile strength is accounted for by the least squares regression equation obtained.

As a point of interest, it might be mentioned that the solution of the above problem required 32 hours punch-card machine time and about 4 hours on a desk calculator.

We would like to thank Professor J. D. Bankier of McMaster University for recommending Hoel's procedure to us and for his advice in adapting the procedure to punch-cards.

REFERENCES

1. Hoel, P.G., Introduction to Mathematical Statistics, Wiley, New York, 1949, pp. 110-120.

2. Lowe, John "Solution of Simultaneous Linear Algebraic Equations Using the IBM Type 604 Electronic Calculating Punch," Proceedings Computation Seminar, December,1949, International Business Machines, New York, 1951, pp.54-56.

GAGE CONTROL

C. H. Borneman
General Electric Company

Quality control usually requires measurement of the product. Mechanical measurements usually require gages. It is not economical to make gages perfect nor to keep them perfect. Gages must, therefore be controlled as to their own quality.

A gage control system is necessary to assure the accuracy of gages that are in use in the shop; or in the tool crib or gage crib ready for reuse; or that are in reserve stock. A gage control system must also assure the accuracy of the master gages that are used to check other gages.

The following can cause gage sizes to change:

Wear of the surfaces that contact the work
Change of adjustment of adjustable gages
Springing of gage frames beyond their elastic limit

If proper heat treatment and aging practice has not been used, gages can also change size because of molecular change of the gage material - usually in the direction of growth-increasing the gage size.

A gage control system should include the whole system of requesting, ordering and inspecting new gages; reinspecting gages periodically during their useful life; scheduling the reinspection dates; systematizing the return of gages for inspection; arranging for the repair of salvable gages; recording gages according to kind and size of gage or kind of product being inspected; recording the location of gages both when not in use and when in use; and recording the final disposition of the gages when they are not economically salvable, or when they are obsolete, or the piece parts which they measured are no longer being made.

Three gage control systems are in use. They are:

1. The required return of gages at the end of each job or shift, or other fixed period of time.
2. Gage reinspection at the point of use.
3. The date marking of gages indicating the limit of time beyond which each kind of gage should not be used until its accuracy has been determined by reinspection.

For intermittently used gages a relatively safe system is to require that the gages be turned in at the end of the job or at the end of the shift or day or week, and that the gages be reinspected before reissue for use.

Gages which are in constant use are sometimes inspected at the point of use by a traveling gage inspector, or they are picked up on the job and exchanged for newly-inspected gages on a regular schedule. This system requires identification numbers on the gages and a careful card record system which must be conscientiously followed.

The interval of time between reinspections for a gage should be based on experience for each type of gage and the kind of material of which the wear surfaces are composed. This interval of time between inspections should be noted on a master card on which is contained all the other pertinent information relating to the gage, including the kind of gage, e.g., plug or ring or snap; the size; whether the gage is standard or special as to design; and an identifying number. This number may be one of a single consecutive series of numbers; or a number designating the kind of gage and size in the digits to the left of a dash and the serial number for that kind and size of gage to the right of the dash. If the gage is for a special product, this should also be noted on the master card.

Besides the prescheduled date for reinspection or the interval of time between inspections, a further safeguard may be employed to prevent the use of gages beyond their inspection interval limit. This can take the form of a color code in which two or three colored spots are put on the gage. The largest spot denotes the month; the second spot denotes the week; and, where necessary, a third spot denotes the day of the week beyond which the gage should not be used. Manufacturing and inspection personnel must be informed of the meaning of the color spots and must have ready access to reference charts showing the respective color for each month, week of the month, and day of the week. As the color spots on a gage that is being used beyond its limit date will be of different colors than the color spots on the gages that are in legitimate current use, the void gage is readily noticed.

This color code, which is used in a number of General Electric Company plants, has the added benefit of giving assurance of gage accuracy to customers' inspectors who visit or work in the plant because they can readily detect any gage which is being used beyond the date on which it should have been reinspected.

Details as to this color code are as follows:

Colors may be applied with a brush fitted into the cap of the paint bottle or by the fountain pen type of applicator which uses a felt brush in the end of a fountain pen shaped paint container. When the marking must be put on the surfaces of gages which would be subject to handling, a shallow counterbore for each color spot is put into the gage at the time the gage is manufactured.

The paint used in our General Electric Company plants is a paint which is not readily removable by anything except acetone. Sets of bottles or fountain pen markers are assembled into kits. Care must be exercised to be sure that the colors do not change due to settling out of some of the pigments. In filling the bottles, it is necessary that the paint be thoroughly stirred before such filling, and at the time of use the bottle or fountain pen should be shaken to stir up the pigments in the paint.

The colors used are as follows:

	Dykem Cat. Letters		Dykem Cat. Letters
January - White	DW	July - Pink	DXX300
February- Purple	DKG	August - Dk.Green	DRJ
March - Yellow	DLT	September- Dk.Blue	DSL
April - Light Green	DXX294	October - Orange	DNX
May - Red	DNC	November - Brown	DTM
June - Light Blue	DMF	December - Gray	DT

First Week - Brown	Monday - Purple
Second Week- Gray	Tuesday - Dark Blue
Third Week - Yellow	Wednesday- Pink
Fourth Week- White	Thursday - Red
Fifth Week - Black (DB)	Friday - Orange
	Saturday - Green

It must be made clear to all concerned that the color spots represent the end of a period of time and that the gage must not be used beyond the end of that period of time without reinspection. The color spots do not indicate the date of the last inspection of the gage.

When experience shows that the usable period for a given kind of gage should be lengthened or shortened for overall economy, the period length should be changed, but only with the approval of the quality control supervisor or the gage supervisor. If, on inspection, a gage is found to be still usable, but approaching its near limit to the extent that experience indicates the gage could not safely be used through the usual period, then the inspector should shorten the next period to correspond to the remaining wear life in the gage and so color mark the gage for its next inspection. He should also enter the corresponding date in whatever card is used for calling the gage back for reinspection at the end of its usable period.

A policy must be established as to whether the usable periods for all gages should be such that all gages will always last through the period without wearing beyond the permitted gage wear limit. It may be economically sound to set the periods of time between gage inspections such that as many as 10% of the gages, upon reinspection, will be worn beyond their wear limits. Any such percentages of gages permitted to go beyond their wear limits would, of course, vary with different kinds of work and different accuracy tolerances. Very critical work might require that only 5% of the gages could be allowed to go beyond their wear limits during their established use periods. On the other hand, the period may be considered right when as many as 15% of the gages returned for inspection are found to be worn beyond their wear limits. When gages are to be used in different parts of a plant or shop and on different kinds of work - the same gage being available for use under any of these conditions - then the usable time must be set according to the safe wear period of the gage under the worst conditions. Templates and other gages which are subject to very little wear can usually be safely used for periods of three to six months between inspections.

For gage economy and for quality control charting of product dimensions, the indicating types of gages are being used more and more. These may take a variety of forms depending on the kind of measurements being made by the gage and the indicating means used. They may use levers, gears, optical magnification, air flow, air pressure change, electric circuits, or electronics. Most of these gages will measure work that falls within a range of size. When work and gage are in proper relative position the gages are self-adjusting to the work size, i.e., the operator does not have to adjust the gage size to the work size before taking the readings. (The ordinary micrometer does require such adjustment to the work size.) The operator applies the gage to the work or the work to the gage, determines when the gage and work are in the right relative positions, then reads the gage.

Almost all of the indicating types of gages require masters for setting them. Some of these masters are simple plug or ring masters. Gage blocks are sometimes usable as masters. Some masters resemble the work and have all the pertinent dimensions of the work.

Masters also must be measured occasionally to assure their accuracy. The interval of time between inspections depends upon the conditions and the frequency of their use. The danger is that they will last so long because of infrequent use and light contact of the gaging members that it becomes taken for granted that they are eternally perfect. Here again a card record indicating the last inspection of such master gages - flagged for inspection at a given later date - is essential. Expensive contour gages are sometimes accepted with known errors which would be very expensive to correct. These known errors are taken into account in the setting of adjustable gages to the master. The adjustable gage is then set to read the known actual dimension of the master.

Accurate gage measurements require control of the temperature at which the gages are measured. Most gage manufacturers maintain the temperature of their gage inspection rooms to the International Standard of 68°F (20°C). The reason for maintaining a very close temperature range is that all materials expand with a rise in temperature and contract with a drop in temperature. For example, a pin gage 24" in length increases 1/1000" when subjected to a rise of 6½ degrees. If this same gage were measured anywhere else in the world under the same conditions the results would be identical.

One important requirement of an efficient gage room is clean air preferably with the humidity maintained between 57 to 59%. This requirement is necessary for the proper functioning and life of the gage equipment as well as for the comfort of the gage room personnel.

Conditioned air is usually conducted to ducts located in the ceiling near the middle of the room, distributed at low velocity to prevent drafts throughout the room, withdrawn through ducts with controllable registers located in the base boards, and brought back to the filtering and conditioning equipment.

For best temperature control the room should be completely insulated, including the ceiling. Outside walls should be avoided.

Fluorescent lighting is the best type of illumination for such rooms. Heat transfer from the lights should be kept to a minimum.

If natural daylight is used, it should be from the North.

Battleship linoleum of a light tan is usually used as floor and bench covering. The room should be kept scrupulously clean at all times. This pays off in better inspection and in greater respect for gages.

To prevent sudden temperature changes the entrances are protected by vestibules, and, where possible, service in dispensing and receiving gages is rendered through small windows.

The main object of this type of room is to maintain the air at a constant temperature. Large masses of metal, such as surface plates, assist in bringing gages to this temperature if they are allowed to remain in contact with such surfaces long enough. Some gage manufacturers go so far as to put gages that are to be measured in a bath of Toluol to bring them to measuring temperature. For extremely accurate measurements all possible precaution is taken against heat transfer to the gage and the gage inspection equipment from the hands or body of the gage inspector.

MILITARY SPECIFICATIONS AND STANDARDS AS STANDARDIZATION TOOLS— THEIR DEVELOPMENT AND COORDINATION WITH INDUSTRY

Russell A. Moody
Office of Standardization
Defense Supply Management Agency

Standardization, like a cemetery plot, is something the average person never thinks about until an urgent need arises. The cemetery plot can usually be obtained on short notice without any particular disruption of plans. But standardization, on the other hand, requires time, manpower and money intelligently and effectively applied before satisfactory arrangements can be completed. Like the cemetery plot, standardization is something all of us will eventually need, and the need can be depended upon to occur when we least expect it.

Shortly after becoming Chief of the Office of Standardization, Defense Supply Management Agency, where I assumed responsibility for standardizing the items of supply used by the military departments, I bought a new home. These apparently unrelated events, private and public, shared a common problem..... standardization.

The bathtub in my new home required a rubber drain plug so I stopped in at a hardware store to buy one. I learned that rubber bathtub plugs range in size from 1½ to 2½ inches in graduations of 1/8 inch. By the time I drove home, measured my drain, drove back to the store and home again I needed a nice refreshing bath to cool me off.

A domestic inconvenience of this kind seems unimportant, once it's over, until one translates the same lack of standardization into a combat situation where want of a standard part can cost lives, time, and even victory.

Standardization, as all who have studied the sciences know, is nothing new. It began centuries ago with the establishment of systems of measurement by royal edict. It is perhaps regrettable that the visionaries who recognized the need for definite weights and measures could not have solved the problem more scientifically. The distance from the tip of the nose to the end of the thumb of the extended arm of Kind Henry I of England and the average length of the left feet of sixteen churchgoers in Elizabeth's reign, cannot be considered wholly precise criteria for the linear yard or foot. Neither can the total length of three grains of barley corn laid end to end be accepted as a positive inch. But, they were made standard for the whole of England and proved most valuable in blazing a trail to the precision instruments of measurement we know today.

Standardization received its greatest impetus coincident with the Industrial Revolution. Whether mass production methods grew out of or demanded concepts of standardization is a fact buried in antiquity. Mass production and standardization, in any event, were handmaidens of industrial progress.

In America, the most notable contribution to both mass production and standardization originated with Eli Whitney in 1798 and, as is the case with much technological development, it was military necessity which prompted it.

Whitney had contracted to produce muskets for the Government. He conceived the notion that production speed could be increased and the replacement parts problem reduced if parts were interchangeable among muskets. It took him two years to tool-up for the job. He had to devise standards for measurement within his own shop, develop standard templates, and plot and apply standard practices. Just as the Government was about to cancel its contract, muskets began to emerge from the little shop and Whitney's firm convictions regarding interchangeability and standardization were justified.

At the Battle of New Orleans, during the War of 1812, credit for the victory belonged not only to Andrew Jackson and Jean Lafitte, the loyal pirate, but to the fact that the British Army had more stands of arms returned to England for repair and replacement of non-standard parts than it had available for the troops in North America. The direct relation of standardization to military effectiveness had been proved for the first time in American history.

As the United States grew in wealth, area, population and influence, there were parallel expansions in both business and Government. Industrial and political bodies, like any growing thing, are affected by the climate in which they grow. In business, changing conditions required modification of management patterns and eventually bred the modern techniques of scientific management. In Government, the stop and go nature of military requirements, mushrooming during wars and ebbing during peace, resulted in the development of unchecked and unorganized standards and specifications. Frenzied activity to get the items needed to wage war, carried on with technical manpower overtaxed and in short supply, resulted in some specifications and standards which were poorly conceived simply as an expedient to meet individual crises. The net result has been that the military departments have been saddled, over a period of years, with a patch-quilt of multiple duplicating and overlapping specifications and standards.

In the years between World Wars I and II, scattered attempts at standardization were carried on within the War and Navy Departments. There was no cohesive effort within the military until late 1942 when an informal group, the Joint Army-Navy Committee on Specifications, was formed. This body inaugurated the JAN Specifications and Standards system, which was expanded after the Army-Navy Joint Specifications Board was chartered in 1945. Centralization of responsibility for uniform military specifications and standards increased with the Unification Act of 1947 and the establishment of a reorganized Munitions Board. Within the Board a Standards Agency was created to provide policy and procedures for the development of specifications and standards applicable to all three military departments.

All of the foregoing efforts were conducted on a voluntary basis. That is, they were in no way mandatory upon the military departments and depended, in great part, upon departmental willingness to participate and their availability of manpower and funds. Each military segment tended to operate on a basis of tradition and past experience, gathered within its own sphere, and was reluctant to experiment in "Centralization". The results, as you can imagine, were somewhat less than spectacular.

In 1952, the Congress took a critical look at the military supply systems and activities associated with them, such as cataloging, standardization, inspection and packaging. During the hearings on these matters, the Congress became aware that cataloging of items in the military supply systems and the reduction of such items through standardization were mutually supporting. This concept was reflected in Public Law 436, the Defense Cataloging and Standardization Act which became effective July 1, 1952.

The Act created the Defense Supply Management Agency and assigned its Director responsibility and authority, under the Secretary of Defense, for providing an economical, efficient and effective supply management organization within the Department of Defense through the establishment of a single supply cataloging system, the standardization of supplies, and more efficient use of supply testing, inspection, acceptance, and packaging facilities and services. The personnel, facilities and functions of the existing Munitions Board Cataloging Agency and Standards Agency were transferred to the new Agency, which is known alphabetically as DSMA.

DSMA, headed by Joseph W. Fowler, a retired rear admiral whose Navy career had been marked by wide experience in industrial matters, has been organized with three co-equal offices: Office of Cataloging, Office of Inspection and Office of Standardization. The packaging responsibility spelled out in the law is included in my responsibilities as Chief of the Office of Standardization.

Before I get into the substance of the program I want to point out that all of our work is done by the Departments, not by my office. The Army, Navy and Air Force get together under our guide lines and do the standardization and write the standards and specifications. It is only because of their whole-hearted cooperation that we have progressed.

The complexities of military standardization, or of any standardization involving technical items, require organization as a prerequisite for solution. As you well know, before order can be achieved in any system of facts, a method must be developed to sort out the various kinds of things into some rough form of segregation. This sifting and compartmentalization has been one of the primary aims of the Military Specifications and Standards system to achieve standardization. Our method of organization pivots on a clearing out process: simplification.

A necessary preliminary to simplification, however, is the determination of what is actually in the system. This function is conducted for us by the Office of Cataloging through its item identification program. That Office has developed a classification system which establishes groupings of associated items and then subdivides those groups into more definitive classes. A single technical service, bureau or command of the three military departments is given primary responsibility for collecting all of the catalog data for a particular group. The development of the catalog data is conducted under strict patterns to insure that like items are similarly described so that differences and identicality will be readily apparent during the culling out process. Within each department, duplications are eliminated before the item descriptions are forwarded for final comparison with all other items in the group.

During the cataloging process, therefore, the compartmentalization or segregation operation is accomplished as far as items in the supply systems are concerned, and the data is arranged in such a way that screening can eliminate duplicates.

It is after this essential preliminary work that the simplification operation begins. Simplification to us means the elimination of unnecessary variety to provide fewer areas containing fewer items. This provides for ready identification of those areas where standardization will pay the greatest dividends in the shortest time and reduces the great mass of technical analyses which must be conducted to produce valuable and valid standardization results.

Simplification works this way. Those things which have outlived their usefulness are removed from the system, reducing the clutter during the process of review. Next the two's and three's of a kind are sorted, examined, and combined to afford some degree of unity. With the underbrush cleared, we can see and get at the other scrub growth which will hamper the development of first-run timber.

With this process completed, only the productive and useful species stand to develop into suitable standard items, if you will excuse the mixed metaphor. Our engineers, together with representatives of the Army, Navy and Air Force, submit all of the remaining item data to a critical analysis to determine whether a Military Specification or Standard can be developed which will adequately reflect the actual Military needs for a particular group of items. All possible uses of the kind of thing making up the group are checked and we may discover that perhaps 50% of the remaining items within the group will meet all requirements of the military. A Standard will be published listing only those items in the group which are absolutely necessary. In the future only those items will be designed into assemblies of new equipment, limiting the influx of deviations for that kind of thing.

Since a Military Standard represents only the design essentials of items to be used, it is then necessary to write either definitive or performance type specifications to enable a producer to manufacture the items required. The development of these specifications involves the gathering together of all existing military specifications for all departments, technical services, bureaus, and commands and the determination as to how they can best be amalgamated or eliminated to produce a completely workable document to meet the needs of both industry and the purchasing activity of the Government. When a specification has been developed, usually through assignment of custodianship to one of the military activities best qualified in a specific area, it is circulated through the military departments for coordination and emerges as a document mandatory for use in all military purchases of the items described therein. Where there is inter-service disagreement, the Chief, Office of Standardization makes a decision, subject to approval of the Director, Defense Supply Management Agency.

The foregoing, of course, is oversimplification of the simplification process. There are many of us, I can assure you, who wish that it were as easy as it sounds. This business of limiting variety, for instance, is often a poser. In the interest of efficiency and economy, it seems obvious that the fewer items of a kind carried, the better off our military supply systems would be. That is not necessarily true,

however; meeting the requirements of the military is our prime concern, and it is necessary to keep in mind ease of manufacture, relative availability of items in a mobilized economy, conservation of critical materials, a broad industrial base, and any number of other considerations. These factors must be weighed one against the other before our engineers can make an honest determination that item A should be standard and item B eliminated.

Further, as I pointed out earlier, the preparation of new specifications requires the appraisal and cancellation of those already existing so that confusion will be eased rather than multiplied. The most recently coordinated Military Specification must supersede all previous specifications for the same item or group of items; the research required to develop this data involves checking the documents not only of the three military departments and their sub-units, but also those of the Federal Supply Service, General Services Administration. There are thousands of so-called interim or tentative specifications, issued during World War II, which are being used as contracting documents today. These, I must add, are in areas which have not yet come under the eye of the Office of Standardization.

Our policy, which has been enunciated in a Department of Defense Directive, is either to incorporate in the Military specifications system or else to cancel, by July 1, 1955, all existing departmental specifications used by the Army, Navy or Air Force. This task, which we recognize as a mighty one, will clear the path for progress. Once in the Military specifications system, these specifications will be reviewed under the then existing top priority of simplification to discover duplicating and overlapping specifications and effect proper reconciliation.

Thus far, I have dwelt in some detail on what I am inclined to call our salvage operation. It is the business of trying to rectify past action, to organize and codify a hodge-podge of documents which were generated under the pressure of wartime necessity within the guide lines that then existed. Of greater importance than this operation, in my opinion, is the controlled entrance of new items into the supply systems of the military, and the reflection of latest technological developments in the papers which spell out the tangible needs of the military.

These controls and advancements depend upon the development of basic standards, "the building blocks" of technological items. These basics provide us with a reasonable measure of uniformity through the standardization of parts and components. The military departments are bringing about a gradual agreement on the least number of kinds of component parts needed to keep our vast military program going. In the mechanical, electrical, and other fields, we are narrowing the number of choices to those few which will do the job effectively. This will serve to meet our aim of keeping the supply line as streamlined as possible. Each step diminishes the needless deviations thus simplifying the problems of procurement, production, inspection, supply and even interchange of equipment in the field. These inter-service agreements, for each of our standard documents must reflect the honest approval of the men who have to use the item in battle, are then subjected to a rationalization process to enable industry to produce the least number of varieties having the greatest interchangeability. As in the case of specifications, each standard is checked for feasibility, for

producibility, and for availability during mobilization. Then, and only then, is the document issued as a coordinated Military Standard, which will then be mandatory for the design of all new items incorporating the items listed in the standard.

The development of standards, however, is not confined to the items themselves. It is our experience that different ways of doing things in a technical society can cause as much difficulty as variety in items themselves. We have, therefore, a concurrent program involving practice standards which stipulate the ways that certain technical operations shall be carried on throughout the military departments.

This story has been told many times since World War II, but I must retell it at this point because it highlights precisely why we are concerned with practice standards. The Packard Motor Company, during the past war, was called upon to produce Rolls-Royce engines in this country. Because the British drew their plans from a different perspective than ours.... technically, from first angle projection as contrasted to our third angle projection.... two years were spent re-drawing their plans so American industry could produce the British item. You would not be startled, I am sure, if I pointed out that dissimilarity in drawing practices is not limited to confusing international cooperation of this kind but has seriously impaired cooperation even among our own military services and among American industries as well.

We have made considerable progress in the field of drawing room practices, both nationally and internationally, and anticipate much less difficulty when a situation like the one just cited recurs. We are also issuing standards for packaging, packing and preservation practices involving military items. The evolution of this type of document depends in a large part upon the adequacy of our staff, funds, and priority workload, for much of the work in the standardization of practices must originate within our own office.

It would be rewarding, I think, to examine in some detail the ways in which our specifications and standards are developed, aside from the process already described in which cataloging refinement and commodity simplification are reflected.

Primarily, specifications and standards are designed to serve the needs of our fighting men; that is the overriding consideration at all times. It is our national conviction, which has been justified in battle, that the men fighting to support our Republic must be sustained, clothed, fed, armed, protected, and medicated in the best ways known to man. All this must be done while they are being scientifically welded into a potent striking force. We believe that the one makes possible the other.

The Department of Defense must implement this conviction, which it shares, with a constant realization that our means and our resources, including manpower and production, are not inexhaustible. Therefore, all defense plans must be tempered by conservatism.... conservationism might be a more appropriate term. We must protect our manpower first, for we value human life above all else save human liberty, but we must protect manpower with basic items if they will serve, rather than with highly elaborated developments. Our effort is constantly directed

toward finding the lowest possible common denominator among items necessary to support our troops both at home and abroad.

In service, certain items fail to achieve the optimum results. The reports filter back through the field commanders and new developments are undertaken to overcome the deficiencies reported.

At the same time, our scientists, researchers and inventors are devising new ways of making present satisfactory equipment do a better job. Many of our universities are engrossed in solving basic problems in both pure and applied science which have military applications. American industry and ingenuity daily offer new approaches to every conceivable requirement of the Department of Defense on land, in the air, on the sea or under it.

All these needs are answered, by one means or another, and these ideas, combined with all other available information, are screened under the rugged criteria of practicability and applicability. Those which are not feasible are rejected, returned for further study or held for future investigation where warranted.

Those projects which are considered ready to go, are put through their paces after all details are subjected to logical evaluation and all concerned have had a chance to punch holes in the scheme. The item which survives field trials is improved, where necessary, retested, and accepted as a standard item to replace the old one in the military supply system. Since the centralization of policy, planning and coordination of standardization in the Defense Supply Management Agency, all of these development items, from combat ration to complex technical equipments are covered by one or more specifications, standards or drawings coordinated by the services involved.

As an item of supply begins to shape up as a promising part of the system, technicians of the appropriate bureau, technical service or command concern themselves with preparing the descriptions under which they will be bought because the Department of Defense is primarily a consumer rather than a producer of goods.

These documents are drafted under the supervision of one of the military elements most concerned and best qualified. This unit is termed the custodian for that specification or standard and is responsible for preparing a draft which represents the combined views of the three departments. As such documents are prepared, cognizance is given to the incorporation of standards already agreed upon so that the processing, appearance, packaging, unit components, and so on will conform to standard requirements. This process guards against perpetuating the multitude of unnecessary variations which technical ingenuity, when unleashed, can devise. This measure of conformity does not, however, shackle creative talent but rather, realistically provides a sharp reduction in pointless deviations.

Documents emerging from this detailed process are distributed to industry through established channels for industry coordination. This industry appraisal may either be carried on through our industry advisory committees or through our normal channels of supply. The prime purpose is to insure that military demands are realistic in terms of the industrial machines which must meet those demands. If they are too far

out of line with normal industry practice, or are too restrictive in limiting the number of manufacturers who could provide the items, they are corrected.

We recognize that a smooth, efficient flow of goods from multiple producers is the only effective way to support our military machine; specifications must reflect both industrial ability and capacity. If, however, it is the considered judgment of our technical people that the demands must be met to complete a specific military mission, then the industrial machine must be modified to provide a solution to production of that particular item. The military mission is the preeminent consideration in the final analysis; failure to accomplish our military mission may spell finis to our industrial might in a strong, free society.

These "industry-tested" items are finally reduced to a formal written document which reflects the best thinking of both military and industry engineering minds. These specifications, in accordance with law, provide a means for the Government to obtain, upon the widest possible base of open competition, the best material, complying with these documents, at the lowest price consistent with reasonable production costs.

Coordination with industry is not a one-way street. It is a busy thoroughfare with traffic in both directions, usually using a preparing activity as a turn-around point. We use this technique to get the national standards bodies consisting of representative industry to coordinate with us in developing documents, both inside and outside Government, which have a basis in common understanding and do not contradict the best judgment of either cooperating element. Definite strides have been taken in recent months to set up procedures whereby the engineering skill of the United States can best be pooled to make certain that no oversight occurs in the evolution of essential documents for making efficient use of our industrial potential.

I have purposely left discussion of the inspection side of the DSMA program to the last because it will be of greatest interest to this group. Our technical people in the Office of Standardization and their opposite numbers in our companion unit, the Office of Inspection, have become increasingly conscious of the importance of a rational approach to determining the quality of the goods which are procured by the military.

We have only begun to promote the idea that the quality level should be spelled out in the specifications so that a producer will know precisely what kind of performance he is expected to provide in the manufacture of an item. This practice will do away with the necessity for additional instructions and will provide acceptance of uniform quality throughout the military departments, easing one of the headaches which has plagued industry in the past.

Increasing reliance is being placed upon rational sampling plans, on acceptable quality levels, and classifications of defects. A realization of the limitations of time, personnel, and the attendant difficulties of 100% inspection have led us to a thorough inquiry into statistical quality control principles and techniques. Evidence of our growing concern with the quality of the product delivered, and with

military determination of that quality, is evident in the recent decision to remove the Inspection Division from my office and reconstitute it as a separate, co-equal Office along with Cataloging and Standardization. This move guarantees that inspection will have the stature and full-time attention that we know it deserves.

Today, more than ever before in the history of the military, there is a strong movement in the direction of rationalizing our acceptance techniques and our quality assurance provisions. It is my hope that you will have an opportunity to hear from our Office of Inspection at your next annual meeting because I am confident that there will be progress in the next year which will prove both interesting and gratifying to the people in your profession.

This brief statement may have revealed to you that no new development has been overlooked in our desire to secure a maximum return for the minimum effort to get assurance of quality and performance in the field.

All these steps are being taken so that standardization may be achieved realistically and on a sound basis. Our efforts are conscientiously subjected to the scrutiny of our industrial partners on the production line to the end of achieving economy in production, economy in use and economy in expenditure of men, money, and materials in support of our national military mission, which is winning, and keeping, the peace.

COST OF QUALITY

W. H. Lesser
General Electric Company

To most of us engaged in quality control work, the most challenging part of our job is that of selling new ways to solve old problems. Scientific sampling, process control charts, product quality audits and planned experiments are still comparatively new ways of reducing cost and increasing profit. Since an overall quality control program is most effective when integrating and coordinating all of the quality improvement activities in an organization, the quality supervisor must become quickly educated upon his relative position in the organization.

As a new member of the management team, he must understand that his program is one of many services that contributes to manufacturing output and affects profit margin. Management will measure the cost and results of his program, and compare his effect upon profit margin with that of other services, such as: personnel relations, time and motion study, or advertising. In this paper we shall examine several yardsticks of quality cost that enable management to evaluate the performance of a quality control program.

As is true with most technical work, it is so easy to spend time on the new and interesting parts of our job, that we sometimes lose sight of our basic responsibility to management namely - to increase profit. Mr. Paul Hoffman, Past President of the Studebaker Corporation, has said of the responsibilities of business men, [1] "The greatest contribution any business man can make toward stability is to operate his business profitably. Every business man owes this major responsibility to his stockholders, his associates his workers and the public. Only a profitable business can offer security to its employees."

Foremost in the mind of every manager when setting his annual budget is:

> How much will this activity cost, and
> What will I get out of it in terms of
> increased profit, higher inventory
> turnover, cost reduction and improved
> market position?

With the multitude of sales, engineering and manufacturing activities, each clamoring for its share of the budget dollar, the quality control activity must compete on an equal footing by proving simply and without question, value received for services rendered. An ambitious quality supervisor will sometimes attempt to justify his budget request by statements such as:" "Quality is difficult to measure", or "It is a good idea to spend money on quality control", or "We need more inspection around here". Smart management will seldom accept such unfounded statements, but will listen to statements such as: "Our cost of quality for 1952 was

so many dollars", or "This vendor rating program will increase inventory turnover so many %", or, "Operator quality mindedness has improved as shown on this % defective chart". By making specific statements, and of course, be able to justify each one, your product manager will identify you quality activity as paying its own way and not tag is as a candidate for expense reduction.

Although each industry and manufacturing concern differs in its organization for the control of quality, there are common yardsticks of quality costs that can be applied to all companies. These yardsticks for management must:

Measure quality level in dollars,
Reflect quality level changes,
Provide a ready comparison
with profit margin, manufacturing
output, product cost and operating
expense.

It is the responsibility of each quality supervisor to establish reliable, but simple indices of quality cost for each important phase of operation for which he is held accountable. Existing cost reports are usually based upon the traditional classification of accounts, and the quality control engineer must often spend unnecessary time in breaking down product costs into significant workable form. Accountants must be convinced of the necessity to report quality costs by responsibility, cause or effect in order to show the results of corrective action.

Let us now look at the basic elements of quality cost found in any manufacturing organization. Here we see (Chart 1) that these quality costs in any company amount to so many dollars a year, and can be broken down as follows:

Identified Costs - scrap, rework, inspection, test, quality control, and customer complaints

Hidden Costs - production delays caused by defective work, lost business due to a poor quality reputation, inherent production design weakness and extra cost due to poor quality planning

This total cost is equivalent to so many cents for every dollar of sales. In some companies the hidden quality costs far exceed identified costs, and educated judgement is required to arrive at a dollar amount. The importance of determining the amount of quality cost cannot be over emphasized, for management should demand a yardstick to measure performance. Many manufacturing organizations now have a good story on the amount and nature of obvious quality costs. Much more work is necessary to pin down those hidden and elusive quality costs in order to fully realize cost reductions from an integrated quality improvement program.

Let us now look (Chart 2) at the cost of quality as compared to other financial statistics. So many dollars is equivalent to

- % of the output of the - Department, - % of the expenditures for plant and equipment, - % of wages and salaries paid, and - % of dividends to stockholders. For telling your own company quality story, use this kind of comparison, and replase stockholders by profit margin and the - Department by your machine or assembly department, but, always compare quality costs with other items of cost familiar to your own manager.

Now let us see where on the financial statements we can find the amount of quality costs. Here (Chart 3) are extracts from a typical profit and loss statement - the bread and butter report of every product manager. Included in this report are practically all of the costs in which we are interested. Behind each amount shown here are one or more financial reports which give details comprising the total.

The first item is net sales billed. This is the base upon which most financial ratios are calculated, and represents the income from which a manager must pay for his material, labor and services, and still have something left over for taxes and profit. The four basic elements of product manufacturing cost are: direct material, direct labor, overhead and engineering expense. Within these cost elements can be found most of the money upon which a quality control department can work.

Direct material, which is often the largest amount of the four items, is always high in priority on the cost reduction budget. Improved vendor contacts and certification plans, periodic review of specifications and methods with suppliers, adequate incoming material inspection and process controls are typical quality control activities. From these activities we expect measurable results in reduced inventory turnover, reduced scrap and reduced production delays caused by rejected material.

Every organization should establish for itself each year a realistic cost reduction budget, made up of specific projects to reduce operating cost. In this program quality control activities can play a major part by making use of: publicity campaigns, effective process controls, improved job instruction, reduction of scrap and inspection rejects. Such a program, when actively supported by management is bound to improve the performance and quality mindedness of direct labor.

Overhead, sometimes called burden, is the one account where heads quickly roll under the axe of expense reduction. Lets beat our manager to the draw by showing him reductions in test and inspection cost, production delays due to defective work, tool maintenance, idle time and unnecessary services. Activity in this direction calls for cooperative effort from all members of the cost reduction team.

Engineering expense, including customer complaint cost, is a fertile field for statistical quality control. Planned experiments, analysis of data, product pilot runs and tolerance analyses can show savings not only in cost, but also in time when getting anew

product on the market. Considerable attention is given by advertising men to the promotion of product quality. This publicity looks good in print, but we must be sure that our outgoing products on the shipping platform are everything the engineers specify and the salesmen claim.

The next chart will show how % final inspection rejects are translated into dollar cost. Here we see (Chart 4) the amount of money spent on scrap and rework in the assembly of an electrical product. Note the high correlation between % final inspection rejects and dollars scrap and rework. At this point, a drive was made to arrest the increasing reject trend by a coordinated program to improve product design, tools, assembly methods and incoming material condition. Note the results of this program as reflected in % inspection rejects and manufacturing loss reduction. Always express your quality control improvements to management in dollars and cents.

The amount of direct scrap and rework labor caused by defective purchased material is usually easy to obtain from inspection and cost reports, and is a good yardstick to measure the effectiveness of a vendor relations program. Here we see (Chart 5) a breakdown of the amount of loss, including overhead, incurred by defective vendor material and the recovery - by commodity, by source and by type. Often purchased material losses, as well as total losses follow a pattern by which about 10% of the items account for 80% of the loss. A commodity analysis will show which vendors are causing most of the trouble. From this analysis we can determine the kind of investigation and corrective action necessary to obtain the desired results.

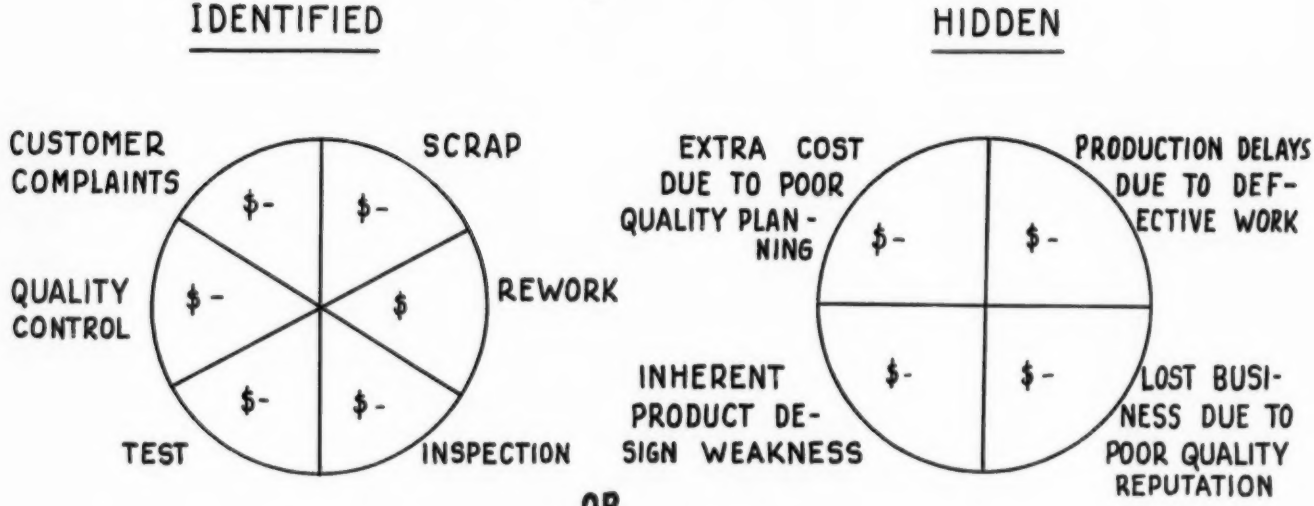
Vendors fall into natural classifications by reputation, size, certification or past performance. Higher recovery is always realized on those vendors selected by the purchasing agent according to the "Delivery, price and quality" formula. Poor recovery is usually obtained on reworked material due to the continual squeeze on the user to immediately repair defective material in order to maintain output. This is where a progressive vendor relations program will save money and reduce inventory by halting defective purchased material before shipment is made to the user.

Your key elements of quality cost can be presented to management in a similar manner. Charts and dollars go a long way to convince your manager that the quality program is self supporting, but always bear in mind that the quality yardstick is only one of many used by top management to measure company performance. The keen quality supervisor displays good business sense to compare his cost objectives with those of other services, also competing for budget dollars. It is only after we display a knowledge of the facts upon which we will be measured for performance, that we can expect a general manager to wholeheartedly support our quality control program.

1 "The Great Challenge to Capitalism" by Paul G. Hoffman, New York Times Magazine, September 8, 1946.

COST OF QUALITY

$ -



OR

-¢ PER $ OF SALES

CHART 1

COST OF QUALITY

$-

EQUIVALENT-TO

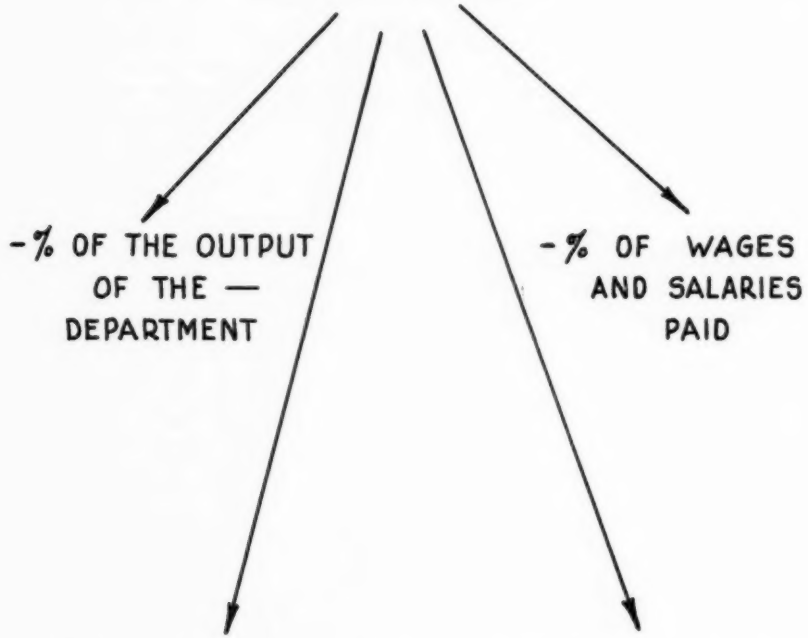

-% OF THE OUTPUT OF THE — DEPARTMENT
-% OF WAGES AND SALARIES PAID
-% OF THE EXPENDITURES FOR PLANT & EQUIPMENT
-% OF DIVIDENDS PAID TO STOCK HOLDERS

CHART 2

PROFIT & LOSS STATEMENT

*"SOURCES OF QUALITY COST"

NET SALES BILLED

Less: COST OF SALES

DIRECT MATERIAL	STANDARD MATERIAL EXCESS OVER STANDARD* SCRAP*
DIRECT LABOR	STANDARD LABOR REWORK* EXTRA COST* STARTING COST*
OVERHEAD	INSPECTION*, QUALITY CONTROL* TRAINING, IDLE TIME PURCHASING*, PRODUCTION*
ENGINEERING	PRODUCTION ENG'R., LABORATORY TESTS*, PILOT RUNS*, CUSTOMER COMPLAINTS*

NORMAL GROSS MARGIN

Less: COMMERCIAL
ADMINISTRATIVE
ACCOUNTING

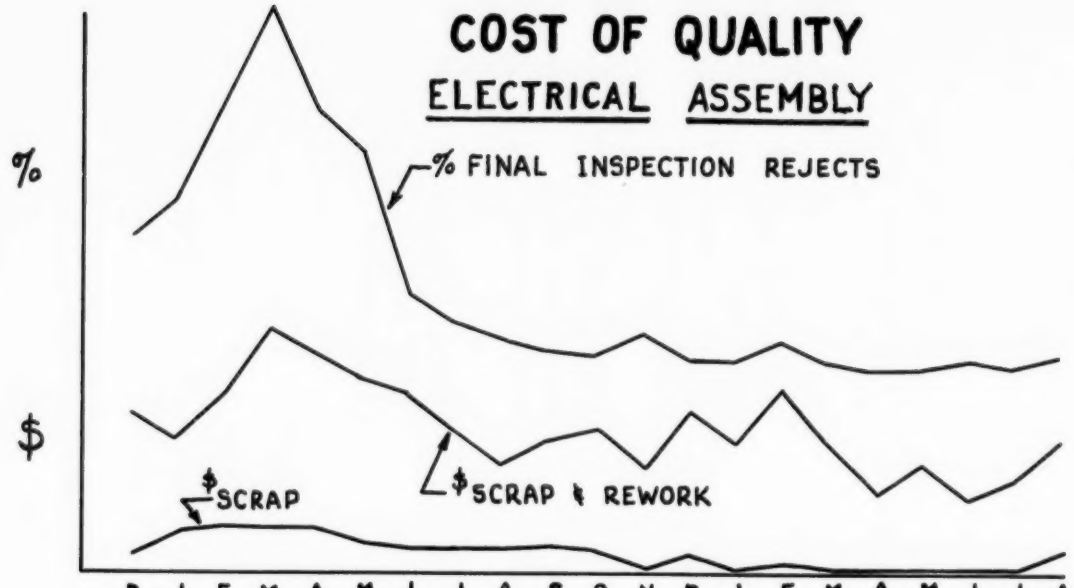
ADVERTISING
CONCESSIONS TO-
CUSTOMERS*

NET INCOME BEFORE TAXES

Less: TAXES

NET PROFIT

CHART 3



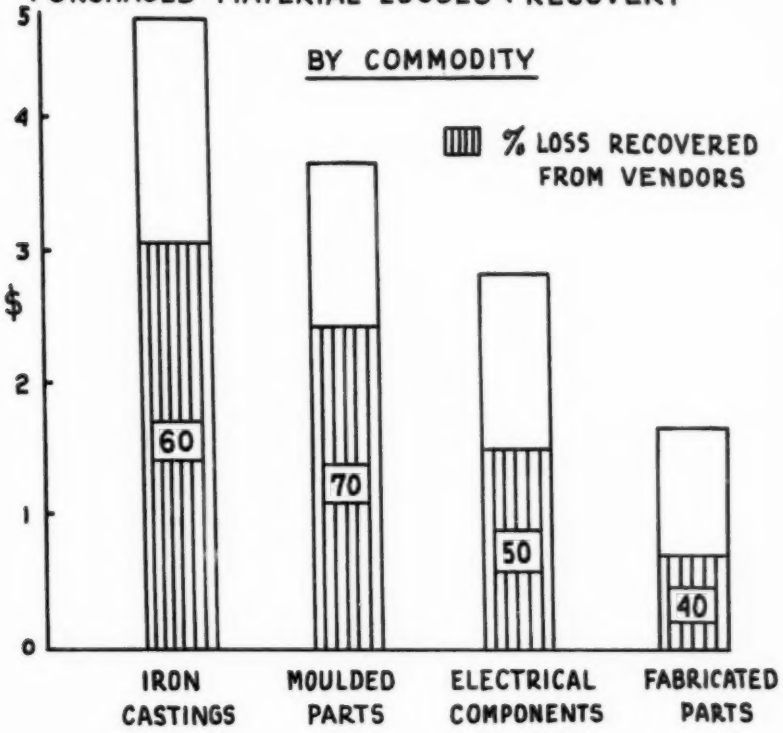
ANNUAL IMPROVEMENT

SCRAP	$ -	75 %
REWORK	-	30 "
% REJECTS	-	60 "

CHART 4

COST OF QUALITY

PURCHASED MATERIAL LOSSES & RECOVERY



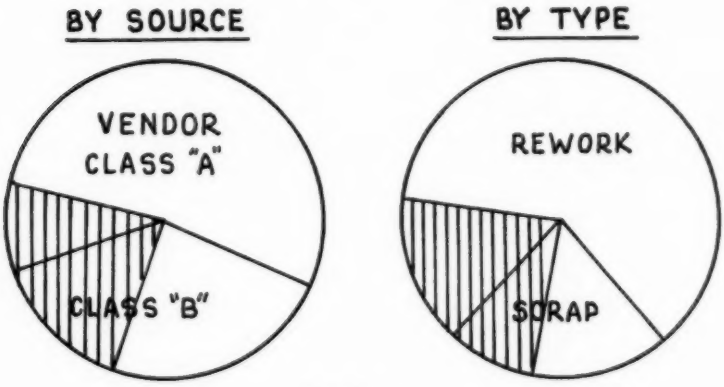


CHART 5

A TRADE UNIONIST LOOKS AT QUALITY CONTROL

William Gomberg
Int'l Ladies' Garment Workers' Union, AFL

A trade unionist's interest in quality control combines the general curiosity of the average scientifically minded citizen in advancing technological progress with a specific interest in its impact on the employer-employee relationship.

Modern statistical quality control techniques impinge upon the employer-employee relationship along the following lines:

1. Statistical quality control introduces another factor of technological change threatening the stability of historical job requirements and relationships in specific industries.
2. Statistical quality control has been offered as one of the criteria in developing standards of performance under otherwise orthodox time wage payment plans or incentive wage payment plans.
3. Statistical quality control has to an unprecedented degree popularized the application of sound statistical theory to hitherto empirical methods of deriving standards of production and expected worker performance.

Let us start first with a discussion of quality control as a factor of technological change. The introduction of statistical quality control has meant a reduction in the number of inspection jobs for workers for a given unit of production. The problems presented to the trade unionist have been less difficult to handle, if anything, than the vast displacement in historical job relationship caused by more dynamic and drastic sources of technological change. The U.S. trade unions have developed a set of protective devises and a code of industrial jurisprudence to cope with these developments. They are designed to encourage a maximum of free technological development in our society and simultaneously protect the interest of the workers affected. In civilized communities today we have ceased to use the bromide -- "you can't stop progress" to conceal the assumption that the workers must bear the entire brunt of hardships that normally must arise in a dynamic expanding society. I hardly think that this is the place to examine these codes and practices closely; there is vast literature on the subject to which reference may be made.

Let us therefore proceed to point number two. A set of expected worker performance underlies most wage incentive payment plans. The measurement of this performance is made up of two dimensions - quantity of performance and quality of performance. Specifically it answers questions like how many rejects can be expected from a worker before he personally can be held responsible to be an "assignable cause of variation" in the manufacture of some specific product. The application of this thinking to the development of wage incentive payment plans is routine and does not justify extensive development at this time.

The most dramatic contribution of statistical quality control

to industrial relations has been in establishing the philosophy basic to a scientific criterion for the establishment of production standards. Quality control, or better yet, the statistical philosophy promulgated by Shewhart in industrial circles has destroyed the fantastic claims made for magic systems of standard data erected by pseudo-investigators whose avowed good intentions were no match for their lack of technical competence. It is at the bottom of a complete revolution in the thinking of industrial engineers in the specialized field of time and motion study.

My own work in this field was published some eight years ago. Since that time scholars like Adam Abruzzi of Columbia University, Gerald Nadler of Washington University in St. Louis and Harold Davidson of Ohio State University have developed the field far beyond the boundaries I had staked out at that time. Very few contributions have been as important as these approaches for the labor movement in developing an attitude toward the proper application of time study techniques to collective bargaining.

It would be fruitful at this time to review some of the thinking that entered into the formulation of a statistical approach to time study. There had been virtually no departure from the original principles of Taylor in developing time study techniques up to the time of the challenge of the statisticians.

Practically every other scientific field honors its pioneers for their heroic explorations into the realms of the unknown. Most of their findings it is understood have been made obsolete by the investigations of subsequent scholars. Moreover, time study resembles theology much more than it does a scientific methodology. Taylor and Gilbreth are deified rather than honored and the contest among time study practitioners seems to be who can claim the greatest orthodoxy rather than the greatest validity.

The elementary principles set down by Taylor in the early 1900's follow:

a. Divide the work of a man performing any job into simple elementary movements.

b. Pick out all useless movements and discard them.

c. Study, one after another, just how each of several skilled workment makes each elementary movement, and with the aid of a stop watch select the quickest and best method of making each elementary movement known in the trade.

d. Describe, record and index each elementary movement, with its proper time, so that it can be quickly found.

e. Study and record the percentage which must be added to the actual working time of a good workman to cover unavoidable delays, interruptions, and minor accidents, etc.

f. Study and record the percentage which must be added to cover the newness of a good workman to a job, the first few times that he does it. (This percentage is quite large on jobs made up of a large number of different elements composing a long sequence infrequently repeated. This factor grows smaller, however, as the work consists of a smaller number of different elements in a sequence that is more frequently repeated.)

g. Study and record the percentage of time that must be allowed for rest, and the intervals at which the rest must be

taken, in order to offset physical fatigue.

h. Add together into various groups such combinations of elementary movements as are frequently used in the same sequence in the trade, and record and index these groups so that they can be readily found.

i. From these several records, it is comparatively easy to select the proper series of motions which should be used by a workman in making any particular article, and by summing up the times of these movements, and adding proper percentage allowances, to find the proper time for doing almost any class of work.

j. The analysis of a piece of work into its elements almost always reveals the fact that many of the conditions surrounding and accompanying the work are defective; for instance, that improper tools are used, that the machines used in connection with it need perfecting, that the sanitary conditions are bad, etc. And knowledge so obtained leads frequently to constructive work of a high order, to the standardization of tools and conditions, to the invention of superior methods and machines.

A glance over the ten steps reveals the subtle acceptance of a purely mechanical view of nature. Here the substance, the mechanical concepts with which science had been made to work in the field of nineteenth-century physics, was artificially transplanted into the field of time study techniques. The fact that the only claim that could be made for these assumptions was that they were subject to tests of validity by the scientific method was confused with the superficial resemblance between the directions to divide a man's work into simple elementary motions and the appearance of the atomic and molecular theory of physical science. Taylor's paragraph "i", that the best elements be combined into a synthetic pattern arithmetically, completely overlooks the possibility that these atomic movements might combine - as hydrogen and oxygen do - to form a completely new compound - water - bearing little or no resemblance to the original constituents from which it was formed, rather than make up an additive set.

Shewhart's concepts were used to challenge this system of thinking. It was used to point out that the development of scientific production standards must depend upon the derivation of a sound statistical sample from a distribution of work times. These work times if they are to yield a sound, meaningful, statistical sample must, in turn, come from a statistically stable parent population.

It is axiomatic that a statistically stable population must be free of any dominant assignable cause of variation. The working environment making up the surroundings within which the worker uses his tools is subject to so many disturbances that it is well-nigh impossible to list them, much less isolate their individual effects. These variations may be classified as follows:

1. The purely mechanical
2. The psychological
3. The physiological
4. The sociological

The wage system and method wage payment would fall into category four--that is, the sociological. It perhaps is the most dominant influence in this category influencing the amount of work that can be expected. It is therefore sheer nonsense to talk about measuring a

fair day's work independently of the wages or wage system.

The fruitful application of quality control philosophy is perhaps best illustrated by three time study questions proposed by Adam Abruzzi:

1. What are the criteria which may be used to determine that a process is standardized and ready for a time study?
2. What constitute meaningful summarizing statistics of time study data?
3. How valid are current standard data procedures?

The philosophy of Abruzzi's approach to the analysis of time study data may be summarized as follows: "Subjective procedures of rating current productive rates to predict the time of a theoretical normal worker has no function in an objective methodology."

This of course means two things in terms of union policy:

1. Subjective rating procedures are rejected as objective determinants of a so-called scientific production standard in collective bargaining.
2. So-called scientific studies such as the rating project conducted under the auspices of the Society for the Advancement of Management are rejected. The distribution of the subjective prejudices of the industrial engineers of America about what they would like to be a normal working speed is an interesting game but hardly a serious objective investigation.

Let us now talk about standardization of the job. When is the job ready for a time study? An examination of the literature discloses qualitative descriptions that vary from one extreme to another. One authority demands virtually the performance of a micro-motion pattern that satisfies the micromotionist's concept of the most economical combination of elemental motions. Others, like Carrol, are all for taking the study on the job just as it is found. They feel that the job pattern can never satisfy the demands of the perfectionists and tend to the other extreme.

Abruzzi departs from these qualitative descriptions and merely asks one question: Do the performance times of the workers satisfy some rational operational criteria of statistical stability? He has defined this stability on two levels, local stability and grand stability. Local statistical stability is measured in terms of the variation in production times over a continuous series of items made over a few hours or at most a continuous day. Grand statistical stability is measured in terms of the variation in production rates of a series of small samples taken over an extended time interval from increments considered qualitatively to be taken under essentially the same conditions.

The criteria for local stability consisted of building up Shewhart control charts for the means and ranges from as many as 70 subsamples, consisting of three items each, from a total of 210 continuous readings on a single individual. This was done for individual elements as well as overall cycles.

In addition, Abruzzi has introduced a new criterion of stability, the mean square successive difference ratio test.

Significantly small ratios indicate that successive observations are positively correlated; significantly large values indicate that successive observations are negatively correlated. Thus this statistic becomes valuable to detect any trends in the data. The same methods used to determine local stability were in turn used to detect grant stability except that in this case the collection of data took place in very much the same manner that data collection for ratio delay studies takes place. Random readings in continuous series of five items were taken. They were distributed during the four basic work periods of the day. Again, these tests disclosed the existence of grand statistical stability.

$$\frac{\delta^2}{s^2} = \frac{\sum_{i=1}^{N-1} (X_{i+1} - X_i)^2 / (N-1)}{\sum_{i=1}^{N} (X_i - \bar{X})^2 / (N-1)}$$

X_i = ANY READING $(i = 1, 2, \ldots, N)$

N = NUMBER OF OBSERVATIONS

The disclosure of the existence of this grand stability does prove that scientific time study is possible. That is, time study which is performed in accordance with the requirements of sound statistical inference. However, what remains to be disclosed is whether or not the limits of variation for both means and ranges are so wide that the results are of little economic significance for setting production standards. Abruzzi concludes from his findings that control charts with 3 sigma limits for groups of workers show equal degrees of variability even though they show widely different mean unit rates of productivity. The most important finding, however, is that local stability does not necessarily imply the existence of grand stability. They are two independent entities.

Inasmuch as trade unions are interested in the long term characteristics of earning opportunities it follows at once that they are much more interested in the conclusions derived from grand stability rather than from local stability. This, of course, means a change in the conventional method of collecting time study data. Where before a continuous short run study was considered adequate, now small runs of five readings taken at random intervals over an adequate period yield more adequate data.

The next question which we wish to examine is how valid is current standard data practice. Abruzzi examines both the macroscopic type of data and the microscopic data. In the course of his examination of microscopic data he describes the fundamental fallacies of the validity test proposed by Maynard, Schwab and Stegmerton for methods time measurement. I have published this analysis in my review of their book for the Cornell Industrial Relations Review and there is little point in repeating the conclusions here.

Abruzzi's principal contribution to this field arises from his

examination of what constitutes a logical subdivision of a total work cycle into elements. His statistical analysis was designed to determine whether or not the usual subdivisions of cycles from the most macroscopic to the most microscopic were statistically independent. These elements would have to be statistically independent if they are to be used as an additive set; that is if they are to carry a time assignment independent of the element that precedes or follows the element being measured. If statistical independence cannot be established for these elements then a minimum condition for their use must be at least a constant relationship or correlation. By the use of rather complex statistical multivariate analysis techniques, Abruzzi found that for some specific operations the nature of the relationship among macroscopic elements, let alone microsopic elements, was so complex that it varied from operator to operator. In fact, at times it varied at different times for the same operator. (These findings were subsequently confirmed independently by Prof. Gerald Nadler of Washington University, St. Louis, Missouri.) This is hardly the sort of foundation upon which objective systems of microscopic data can be built. These conclusions would tend to substantiate the findings of the gestalt psychologist and industrial engineers like Gillespie and Mundel. "The whole is greater than the sum of its parts."

Does this mean that trade unions will reject all systems of standard data out of hand? Not at all. We are willing to recognize that they are a rather poor empirical attempt to develop a modus vivendi upon which both management and labor can reach agreement. However, under no circumstances would we permit ourselves to be bound solely by the technology of these systems in dispute cases. They at best stabilize a human relationship. They are useful provided we recognize that the failure to achieve a rate developed under standard data does not indict the worker; the rate may very well point up the limitations of the technique.

If the philosophy of the quality control technique did nothing else for the labor movement except undermine the nonsensical foundation of fads, like methods time measurement, that in itself would have made it a major contribution to stable relationships between employers and labor.

THE SPAN PLAN -- A NEW METHOD OF PROCESS CAPABILITY ANALYSIS

Leonard A. Seder and David Cowan
Gillette Safety Razor Company

If any one single expression could be used to characterize the statistical methodology applied in quality control, it would probably be the words "process capability analysis". The use of the frequency distribution and the control chart to study process behavior has led to the concept that every industrial process has an inherent variability that must be tolerated and that any variability beyond this minimal amount may be allocated to a specific cause or causes and, therefore, may be eliminated if it is economical to do so. This minimal amount is often called the process capability. Being measurable, the process capability is a valuable piece of information, since knowledge of it leads to rational action on the process, the specifications or the inspection methods and hence points the way to lowered losses.

This idea is an extremely useful one. It can be explained and demonstrated to the shopman and serves to answer his oft-asked question as to what variability the process may be expected to achieve. In the broader sense, capability analysis is the means of judging the point at which the balance is reached between the cost of quality and the value of quality. The elimination of the extraneous sources of variation and the attainment of intrinsic capability in everyday practice is usually accomplished at less cost than the savings it creates, but the effort at further reduction below the inherent variability usually requires a greater investment than the yield it promises.

With such rich stakes to be won, it might be supposed that the capability concept would be fully understood by quality control practitioners and that standard procedures would be available for its measurement and interpretation. Unfortunately, however, this does not seem to be the case. There are those who consider the capability to be the spread of a frequency distribution of a sample drawn at random (vaguely) from the process. Others insist that a control chart must show statistical control before the capability can be said to have been measured. There is, however, no standard method of applying the control chart specifically for this purpose. Moreover, there is some debate on whether capability should include the total spread in all the streams of product from, say, a 24-cavity mold. One of the authors[1] has emphasized that variability "within-piece", such as out-of-roundness, eccentricity, taper, etc. is often reducible and hence this should not be considered as part of the "irreducible minimum" variability. Neither is it clear whether error of measurement is included in the capability. More advance analysts have pointed out that determination of capability is not complete without an analysis of variance since the latter frequently indicates the possibilities of reductions below the level that the simpler analysis would have called "minimal".

Out of this confusion, the idea is gradually growing that capability is a "molecular" rather than "atomic" aggregation and that it may be regarded as made up of a number of elemental variables, almost all of which can be identified and measured, if necessary. Whether identification is necessary in any specific case depends on whether the process is making defectives; when the capability reaches the point that it can comfortably meet the required specifications - either because of the identification

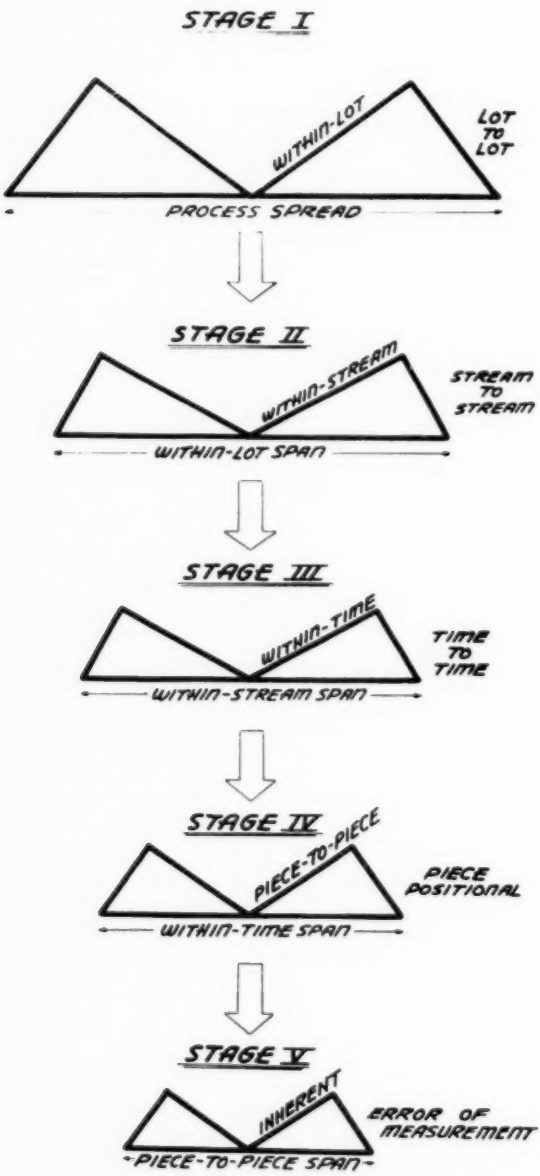


FIG. 1. THE FIVE STAGES OF DISSECTION

and removal of the major elements of variability or in spite of their continued existence - then further subdivision of the capability is unnecessary and, in fact, is economically unsound. Until that point is reached, the problem is not only the measurement of capability but the broader task of attaining the necessary capability. Only when a harmonious relationship between the specified product limits and the process capability has been established can there be assurance that defects will be prevented.

It is with this philosophy and with the hope of clarifying the meaning of the capability concept that the authors have attempted to evolve a standard method of evaluating and describing process capability. This paper is a tentative draft of that method, herein called "SPAN - a Systematic Procedure for Attaining Necessary Capability".

Nature of the Method

Certain of the basic concepts of the SPAN capability method need explanation before the details can logically be presented. The idea of a natural or inherent variability associated with the process is still retained, but it is regarded strictly as the "instantaneous error of reproducibility" of the process. It is, for example, the degree of non-uniformity of five successive pieces in an uninterrupted flow from a single-spindle screw machine, not including out-of-roundness or other variation within a piece. It is the variability found at the same position on each piece of five successive pieces in the steady stream from one cavity of a casting mold or from one die of a punch-press. Note that taper, parallelism, flatness, etc. are excluded from this definition since it is specified that measurements are "at the same position on each piece".

The capability thus defined is seldom attained in actual practice; numerous sources of variation tend to enlarge the extent of non-uniformity found in the product by the time large quantities have been produced under the normal conditions of mass production. Although admittedly numerous, these "extraneous" sources of inflated variability are not mysterious. They can be classified into these broad types:

1. The effects of time, during which materials change, tools wear, operators adjust the machine, solutions become depleted or contaminated, work shifts rotate, etc.
2. The effects of multiple equipment or multiple operation, creating several different parallel paths or streams of product through the process. Examples: duplicate assembly lines, batteries of machines performing the same operation, multi-spindle machines, multi-cavity molds.
3. The effects of tooling, in which the shape, dimensions or alignment are imperfect. Examples: an egg-shaped drawing die, a cocked punch and die-set, a tapered mandrel.
4. The effects of precision of measurement, which necessarily enters into the act of measurement of the product to determine its variability.

To a greater or lesser extent, each of these factors contributes to the total variability in almost every process. The real problem is to discover which of them is making the major contribution in the process at hand, so that appropriate corrective measures can be taken to reduce its contribution, if necessary. The solution lies in isolating and measuring the contribution of these broad factors to the total process variability

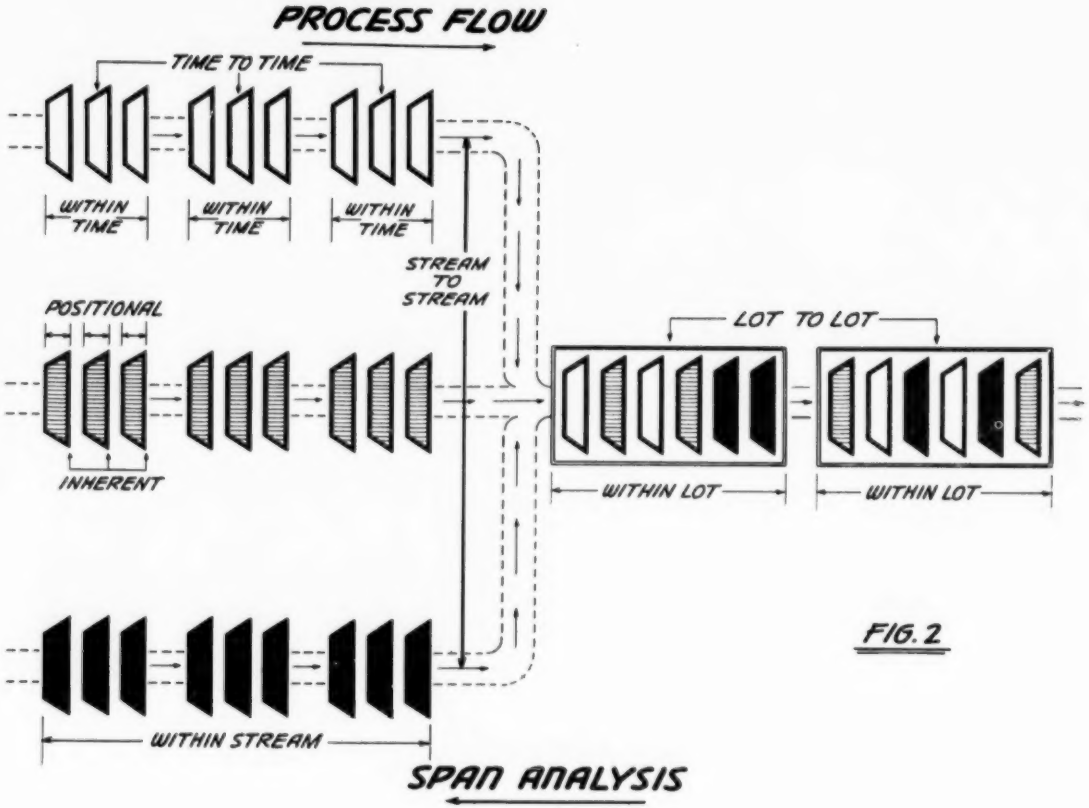

PROCESS FLOW
TIME TO TIME
WITHIN TIME
WITHIN TIME
WITHIN TIME
STREAM TO STREAM
POSITIONAL
INHERENT
LOT TO LOT
WITHIN LOT
WITHIN LOT
WITHIN STREAM
SPAN ANALYSIS

FIG. 2

Thus far, the techniques proposed for accomplishing this quantification have certain limitations. The multi-vari chart,[2] for example, is a "rough and ready" tool to portray the relative magnitudes of the several factors, but its interpretation depends on visual estimation and it ignores the question of statistical significance. On the other hand, the analysis of variance gives much more precise and statistically valid answers, but its use requires considerable training, extensive pre-planning and laborious computations. The object of the development of the SPAN method has been to approach the simplicity of the multi-vari chart while enjoying the precision of the analysis of variance. In addition, the method has an advantage possessed by neither of its forbears in that the amount of data to be collected and the degree of refinement in analysis is governed entirely by the criterion: "is this step needed to attain adequate capability?"

Stages of the Analysis

In outline, the method consists of a sequence of five "stages" or levels of analysis, each representing a more refined dissection of the capability than the one preceding it. The later stages are reached only when the earlier stages fail to reveal a source of variation which, when eliminated or sharply reduced, succeeds in achieving the necessary capability. Thus, the use of the method on any particular problem proceeds only to the stage necessary for its solution.

At each of the aforementioned stages, there are a set of specific procedures which detail the method of sampling, size of sample, method of computing statistical measures and method of portraying the results. A series of standardized tests applied to these results serves as the means of interpretation. These procedures and tests are the same for all five stages of the analysis, save only for appropriate changes in the location and the method of drawing the sample.

In the first stage, the product is sampled as it emerges from the process in lots and an estimate is obtained of the Span[3] of the process. If this is adequate to meet the tolerance requirements, no further effort is expended. If it is not, then the magnitudes of the "lot-to-lot" and "within-lot" components of the Span are estimated from the same data and examined to discover whether elimination of the "lot-to-lot" factor would enable the resulting Span to meet the tolerances. If so, the analysis terminates and every effort is concentrated on discovering and removing the causes of lot-to-lot non-uniformity. If not, the analysis proceeds to the second stage.

In the second stage, the sampling point is moved back so as to enter the individual streams before they are merged into lots, thus dissecting the "within-lot" residual of Stage I into "stream-to-stream" and "within-stream" components. Again the estimates are made and tests applied in the same manner and the decision reached as to whether eliminating the causes of "stream-to-stream" variability will attain the necessary Span.

If such is not the case, one proceeds to Stage III in which the "within-stream" is dissected into "time-to-time" and "within-time" factors. If needed, Stage IV dissects "within-time" into "positional on a piece" and "piece-to-piece". Stage V separates the "piece-to-piece" into "error of measurement" and "inherent". Figure 1 is a schematic diagram of the five-stage procedure as described. Figure 2 shows how the SPAN PLAN analysis proceeds in the opposite direction to the process flow.

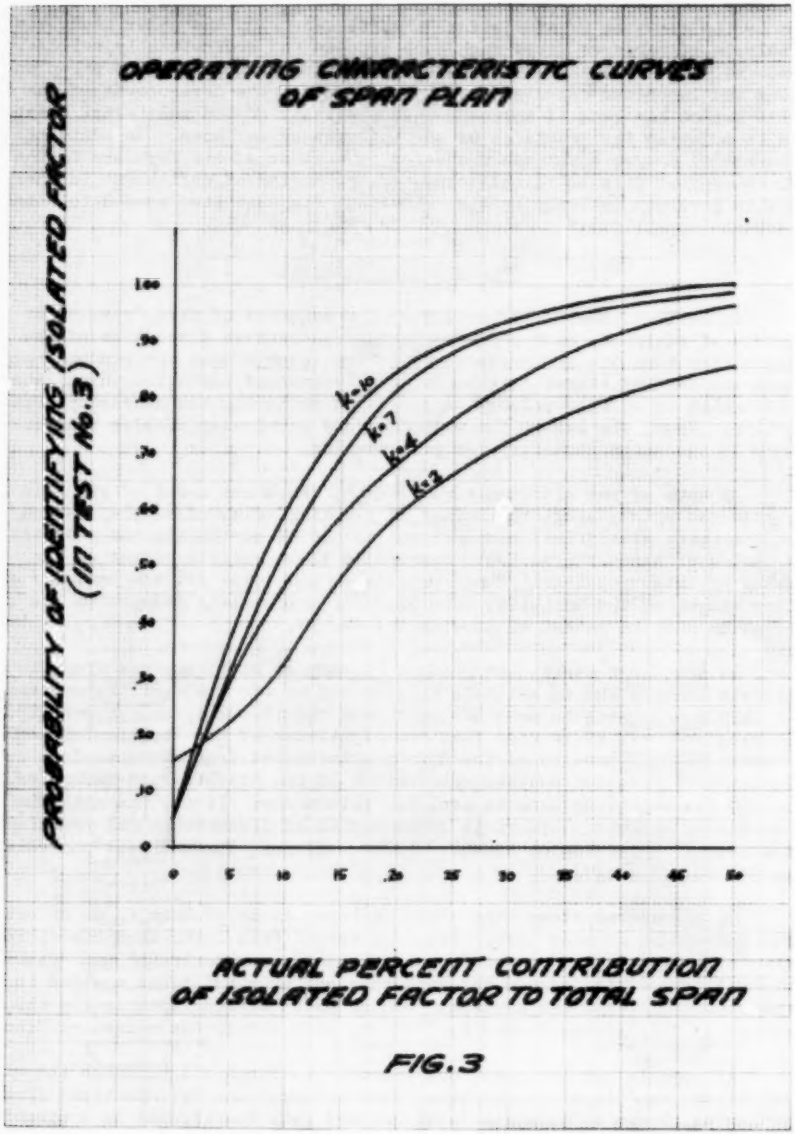

OPERATING CHARACTERISTIC CURVES
OF SPAN PLAN
PROBABILITY OF IDENTIFYING ISOLATED FACTOR
(IN TEST No. 3)
k=10
k=7
k=4
k=2
0
5
10
15
20
25
30
35
40
45
50
ACTUAL PERCENT CONTRIBUTION
OF ISOLATED FACTOR TO TOTAL SPAN

FIG. 3

Several unique features are built into the procedure. They are mentioned briefly here merely to describe the method. The reader who is interested in a more thorough treatment is referred to the Appendix and to the references therein contained.

1. Simple Statistical Measures. Although the technique at each stage is, in fact, an analysis of variance, the only statistical measures employed throughout are averages and ranges. This is made possible by recent work by H. A. David[4] and H. O. Hartley[5] who have worked out the details of using range for the analysis of variance.

2. "Practical" Rather Than "Statistical" Significance. The identification of a source of variation as being worth reducing is based on its actual magnitude rather than its statistical significance. This relieves the user from the dilemma of whether to pursue an effect which is statistically significant but which does not contribute enough to the capability to be of practical importance. Nevertheless, statistical significance has not been over-looked. Rather it has been predetermined in the selection of the sample size required at each stage. In this manner, the user is guaranteed (within certain risks) that any effect which is revealed as practically significant will automatically be statistically significant. The risks are indicated in Fig. 3 which depicts the Operating Characteristic Curves of the SPAN PLAN for detecting sources of variation which contribute varying amounts to the Span..

3. Graphic Portrayal of Analysis. Considerable sales value is placed on the presentation of results in graphic form. To represent the magnitude of the Span at any stage and the relative magnitude of the factors which comprise it, the Pictogram[3] is employed. This device uses the right triangular properties of the root-mean-square to portray the magnitudes of the component and resultant variabilities. In Figure 1, for example, the Pictogram for Stage I shows that the "base line" (6σ) of a frequency distribution of random individual pieces from various lots of product has a length represented by the distance labelled "process spread". Moreover, the mean of the distribution is at the indicated mid-point of this line. So much could have been obtained by the usual method of drawing a histogram or frequency curve of the data; however, the other two lines which form the legs of a right triangle give valuable additional information. The "within-lot" line represents the 3-sigma value obtained by computing a standard deviation for each lot separately and averaging these together; it is thus the variability without regard to differences in lot means. The "lot-to-lot" line represents the 3-sigma value of a distribution of lot means. Thus, the two lines indicate clearly how much of the process spread is attributable to variability within individual lots and how much to differences in means from one lot to another.

Illustrations of Span Plan Analysis

The following several pages give a detailed presentation of the SPAN PLAN with all of the necessary procedures and tables. However, in operation, use of the plan has been simplified to the point where all data, sample identification, calculations, pictogram and analysis are recorded on a single work sheet (for each stage). A sample work sheet is shown in Figure 4.

SPAN PLAN RECORD

STAGE

DATA

MEASUREMENTS	SAMPLE No. AND IDENTIFICATION 1	2	3	4	5	6	7	8	9	10	TOTALS OF ROWS
	LOT NUMBER 20	23	32	45	47	69	73	86	126	127	
	3.5	4.0	4.1	2.0	3.5	2.8	3.0	4.5	4.5	4.0	35.9
	3.7	3.5	3.3	2.9	2.5	4.4	3.0	3.9	3.2	4.2	34.6
	4.7	4.0	3.3	2.0	4.7	3.5	3.8	3.7	2.8	3.0	35.0
	3.-	3.5	4.0	5.0	2.7	2.9	2.1	3.7	3.6	3.6	34.6
	2.5	2.9	3.0	3.5	3.5	4.2	3.5	4.5	3.0	3.6	34.2
	4.0	3.1	3.8	3.0	3.9	3.6	2.8	3.5	4.2	3.8	35.7
TOTALS OF COLUMNS	21.4	21.0	21.5	18.4	20.8	21.4	18.2	23.8	21.3	22.2	ΣX 210.0
RANGES WITHIN COLUMNS	1.7	1.1	1.1	3.0	2.7	1.6	1.7	1.0	1.7	1.7	ΣRun 16.8 / C_{un} 8.50
RANGE OF COLUMN TOTALS											R 5.6 / C_4 2.60 2.15

DATE August, 1952
PART #1105 Razor
OPERATION Assembly
TOLERANCES UTL 5.00 LTL 2.00
Characteristic: Length of end
UNITS OF MEAS. .001" 1.500"
SAMPLING PT. After assembly
10 lots randomly selected from monthly production
POSITION OF MEASUREMENT Random among six positions
ISOLATED FACTOR Lot-to-lot
UNRESOLVED FACTOR

SUMMARY	
$\bar{X}$	3.50
$\bar{R}_{Run}$	2.00
$3\sigma_{us}$	1.97

$\bar{X}$ 210.0 / n 60 = 3.50
RANGE OF Run = 2.00
$3\sigma_{us}$ = 1.97

COMPUTATIONS

	$3\sigma_f$	$3\sigma_{sp}$
	R_f/C_4 = 2.15	$(3\sigma_f)^2$ = 12.53
	$(R_f/C_f)^2$ = 4.6354	$(3\sigma_{us})^2$ = 3.8809
	$(3\sigma_{us})^2$ = 3.8809	Sum Sq. = 4.0062
Diff	= .7545	$\sqrt{\text{Sum Sq.}}$ = 2.00 → $3\sigma_{sp}$ 2.00
$\sqrt{\text{Diff}}$	= .868	$\bar{\bar{X}}$ = 3.50
$\sqrt{n}$	= 2.45	$\bar{\bar{X}} + 3\sigma_{sp}$ = 5.50 → UPL 5.50
	$(3\sigma_f$ if negative)	$\bar{\bar{X}} - 3\sigma_{sp}$ = 1.50 → LPL 1.50
$\sqrt{\text{Diff}}/\sqrt{n}$	= .354 $3\sigma_f$	→ $3\sigma_f$ .354

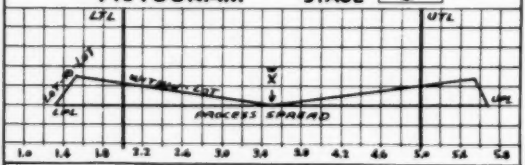
TABLE OF CONVERSION FACTORS

K	2	3	4	5	6	7	8	9	10
n	10	10	9	8	7	7	6	6	6
C_{un}	2.08	3.12	4.00	4.80	5.46	6.37	6.80	7.65	8.50
C_f	1.49	2.01	2.24	2.34	2.36	2.50	2.42	2.52	2.60
Q	.57	.78	.91	1.01	1.09	1.15	1.22	1.24	1.27

TEST No. 1	YES	NO
UTL = 5.00 LTL = 2.00		
UPL = 5.50 LPL = 1.50		X
TEST No. 2		
UTL - LTL = 3.00		
$6\sigma_{sp} = 2X3\sigma_{sp}$ = 4.00		X
TEST No. 3		
$3\sigma_{sp}$ = 2.00		
$3\sigma_{us}$ = 1.97		
Diff = .03		X
Diff / $3\sigma_{sp}$ = .015		
TEST No. 4		
UTL - LTL = Not		
$6\sigma_{us} = 2X3\sigma_{us}$ = Applicable		
TEST No. 5		
Range or $R_{un} - \bar{R}_{un}$ = 2.00		
Range or $R_{un}/3\sigma_{us}$ = 1.00		X
Significance Value (Q) = 1.27		

PICTOGRAM

STAGE I



CONCLUSIONS: NO LOT DIFFERENCES
CHECK AGAINST PRECEDING STAGE: (TEST No 6)
SP/UN = (UPPER LIMIT / LOWER LIMIT) NOT APPLICABLE
REMARKS: TOTAL SPAN DOES NOT MEET TOLERANCES
DECISION: GO ON TO STAGE II

FIG. 4

All pertinent identification of the sample and method of measurement are recorded in the upper right hand block. The data appear in the upper left-hand block. Computation performed on the raw data consists merely of totalling the columns and the rows, getting a grand total (ΣX), recording the ranges within the columns and their total (ΣR_{un}), entering the range of the column totals (R_{if}) and the range of the within-column ranges ($\bar{R}_{un}$). These values are then divided by appropriate conversion factors from the "table of conversion factors" at left center of sheet to obtain figures for the grand average ($\bar{\bar{X}}$) and the within-column Span ($3\sigma_{un}$).

Calculation of the between-column Span ($3\sigma_{if}$), the Total Span ($3\sigma_{sp}$), and the upper and lower process limits (UPL and LPL) are performed in the two blocks in the middle of the page. This completes all of the necessary computations; final results are entered in the "summary" column as directed by the arrows. The bottom part of the sheet is used to draw the pictogram of the process at the particular stage, to apply the tests and to record conclusions and decisions.

While all of the tests are necessary to give a complete file record of the case, most often the pictograms can readily be analyzed by visual inspection. For example, Figure 4, a Stage I analysis, indicates that the process spread is much too large to fit within the tolerances but that "lot-to-lot" variability is not an important contributor to the process spread. Even if the "lot-to-lot" factor were entirely absent (in which case the "within-lot" line would swing down to the "process spread" line) it is clear that the "within-lot" Span is still too large to meet the tolerance limits. Hence the decision is made to go on to Stage II to dissect the "within-lot" into other components in the hope of finding a large one whose removal will enable the remaining variability to meet the tolerance limits.

Another Stage I analysis of a different process is shown in Figure 5. Here the "lot-to-lot" factor is entirely absent and the pictogram reduces to a straight line. In this case, it will be noted that the process has two faults; not only is the "within-lot" too large, but the process is centered poorly -- the process setting is actually outside the tolerances.

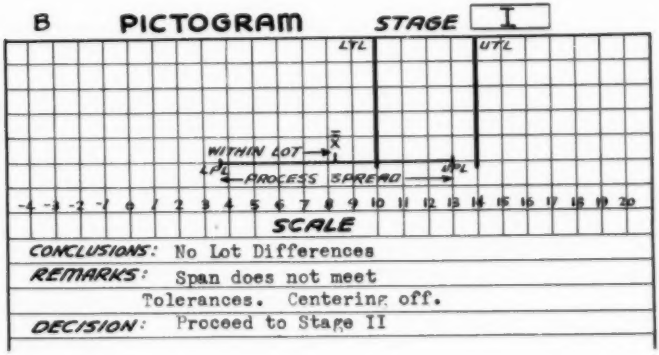


Fig. #5.

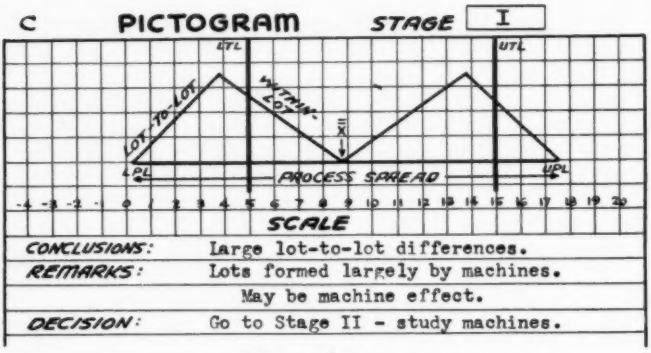


Figure 6.

In the Stage I analysis of Figure 6, the "lot-to-lot" is definitely hindering the process from meeting the specification; its reduction is necessary.

A typical case of a stream effect found at Stage III is shown in Fig. 7. Large difference between machines greatly inflates Total Span of the process

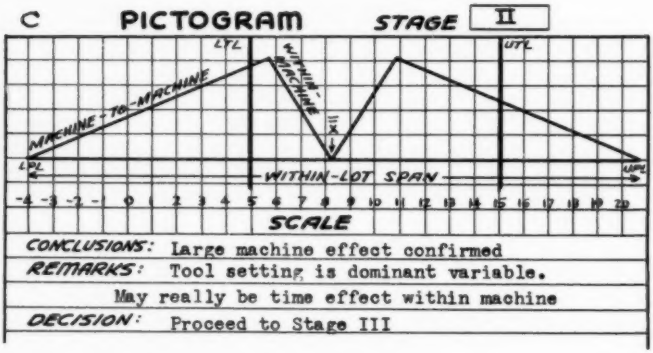


Figure 7.

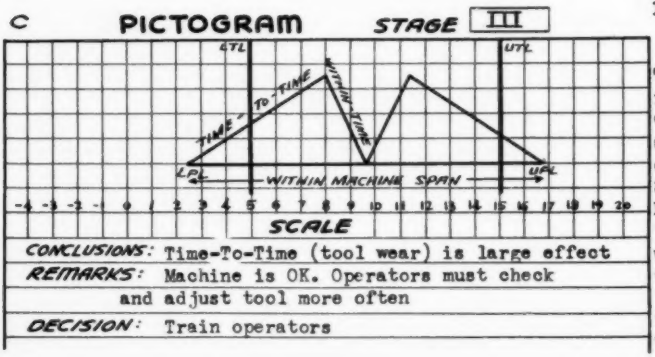


Figure 8.

This is typical indication of existence of short-term time effects at Stage III. Note how easily "within-time" will meet the tolerances when "time-to-time" is brought under control.

Advantages and Limitations of SPAN PLAN

The advantages of the SPAN PLAN may be summarized as follows:

1. It is a carefully spelled-out, comprehensive set of directions for pursuing the capability analysis of almost any industrial process. Although its idiom is in terms of discrete-piece mechanical processes, its principles are equally applicable in the continuous processes of the chemical, textile, paper and other industries.

2. Its procedures and computations may readily be carried out by personnel untrained in statistical methods with a minimum of instruction. It is thus ideally suited for supervised use by persons at the technician and lower levels, and furnishes a means of multiplying the efforts of trained quality control engineers.

3. Because it substitutes an orderly, systematic, statistical method for a haphazard, subjective attack, it makes possible the wider use of better analytical methods at the shop level. If this means that 80% of industry's capability problems can be attacked with <u>better</u> methods instead of only 20% with the <u>best</u> methods, as at present, it is much to be desired.

4. The use of the pictogram as a means of describing and thinking about composite variability provides an appealing approach to those non-statistically-trained people whose cooperation is needed in achieving improved process capability.

The limitations of the SPAN PLAN must also be borne in mind. Some of these are:

1. It is intended as an overall diagnostic method to locate the major sources of variation in a process. It may not necessarily provide all of the information needed for correction. Such information must generally come from a more complete study of the isolated factor, once its importance has been pointed out in the analysis. Suggestions for such supplementary procedures are given in the directions.

2. An understanding of the principles behind the methods is desirable for the supervisor of the investigation, even though the actual work may be done at the technician level.

3. The statistical techniques involve the assumption of normally-distributed quality characteristics. While this is not a seriously limiting feature, it may lead to errors in the case of a quality characteristic having a skewed distribution, for example.

The following pages present a complete draft of the Systematic Procedure.

SYSTEMATIC PROCEDURE FOR ATTAINING NECESSARY CAPABILITY

I. DEFINITIONS OF TERMS USED

The following definitions apply to words with special meaning used in the procedure.

1. Quality Characteristic. The measurable property of a product which the operation or process is intended to produce and which must meet specified tolerance limits. Examples: dimensions, strengths, weights, voltages, torques, etc. The SPAN PLAN must be applied separately to each such characteristic and all of the following definitions refer to any one such characteristic.

2. Process Spread. The total variability in the outgoing product of a process or operation being investigated. It is equivalent to six times the standard deviation of a frequency distribution of a random sample of that product taken over a period of time.

3. Factor. A source of variation which contributes to the Process Spread In the general case, Process Spread is assumed to be composed of the following 6 factors: (1) inherent variability of the machine or operator; (2) error of measurement; (3) piece-positional (i.e. within piece) variability due to tool or fixture error; (4) time-to-time effects; (5) stream-to-stream effects; (6) lot-to-lot differences.

4. Stage. The procedure systematically isolates the contribution of the several factors which make up the Process Spread. This is accomplished in 5 separate steps, each of which progressively subdivides the Spread by isolating one factor. Each of these steps is called a Stage (of dissection and they are numbered I, II, III, IV, V.)

5. Span. The magnitude of variability associated with any single factor or combination of factors. It is equivalent to six times the standard deviation of a frequency distribution of measurements or averages representative of that factor or factors. The Span of any given factor may be thought of as the amount of variability that would be present in the process if no other factors existed.

6. Half-Span. One-half of the Span, or three times the standard deviation, as above.

7. Total Span. The total variability of a process or operation at any designated stage of dissection, equivalent to six times the standard deviation of a frequency distribution representative of that stage. At Stage I, the Total Span is called the Process Spread.

8. Components of Total Span. The two factors which, added together vectorially, produce the Total Span at any stage. For example, at Stage I, the two components of the Total Span are the "lot-to-lot" and "within-lot" factors. Thus, the Total Span is equal to the vectorial sum of the "lot-to-lot" Span and the "within-lot" Span.

9. Isolated Factor. The particular component of Total Span at a given stage which that stage is intended to separate and measure. At Stage I "lot-to-lot" is the isolated factor.

10. Unresolved Factor. The component of Total Span at a given stage which still contains several sources of variation that will be isolated at later stages. For example, the unresolved factor at Stage I is the "within-lot". The component remaining unresolved at Stage V is called the "inherent" variability.

11. Stream. In operations performed on multiple equipment, there exist two or more alternative "paths" which a piece-part may physically follow in proceeding through the operation. Each of these parallel paths is called a stream.

Examples are: multi-spindle machines; multi-cavity molds; duplicate machines; groups of skilled operators on machine, hand or bench work (all performing the same operation).

II. GENERAL PROCEDURE BY STAGES

Stage I.

(The following steps are shown schematically in Figure 9, insofar as they apply generally to every stage of the analysis.

1. Sample a number of lots of outgoing product, as specified in Procedure A, and measure for relevant quality characteristic. It is advantageous to preserve any identifying information pertaining to individual lots, such as time processed, shift, raw material, lot number, etc. for later use if needed.

2. Compute the Average and Total Span, obtain the UPL and LPL, compute the components of Total Span and draw the pictogram, as specified in Procedures B & C.

3. Apply Test #1 to determine whether process is meeting the specified tolerances.

   a. If so, process is considered satisfactory and no further analysis or action is necessary.

   b. If not, proceed to step #4.

4. Apply Test #2 to determine whether process is capable of meeting the specified tolerance range.

   a. If so, the required action is to reset the mean of the process so that it falls at or near the midpoint of the tolerance range; analysis is complete.

   b. If not, proceed to step #5.

5. Apply Test #3 to determine whether the "lot-to-lot" factor contributes 10% or more to the Total Span.

   a. If so, proceed to step #6.

   b. If not, proceed to Stage II. However, it is first necessary to apply Test #5 to learn whether the variabilities within each lot may be considered to be alike, since the procedure at Stage II

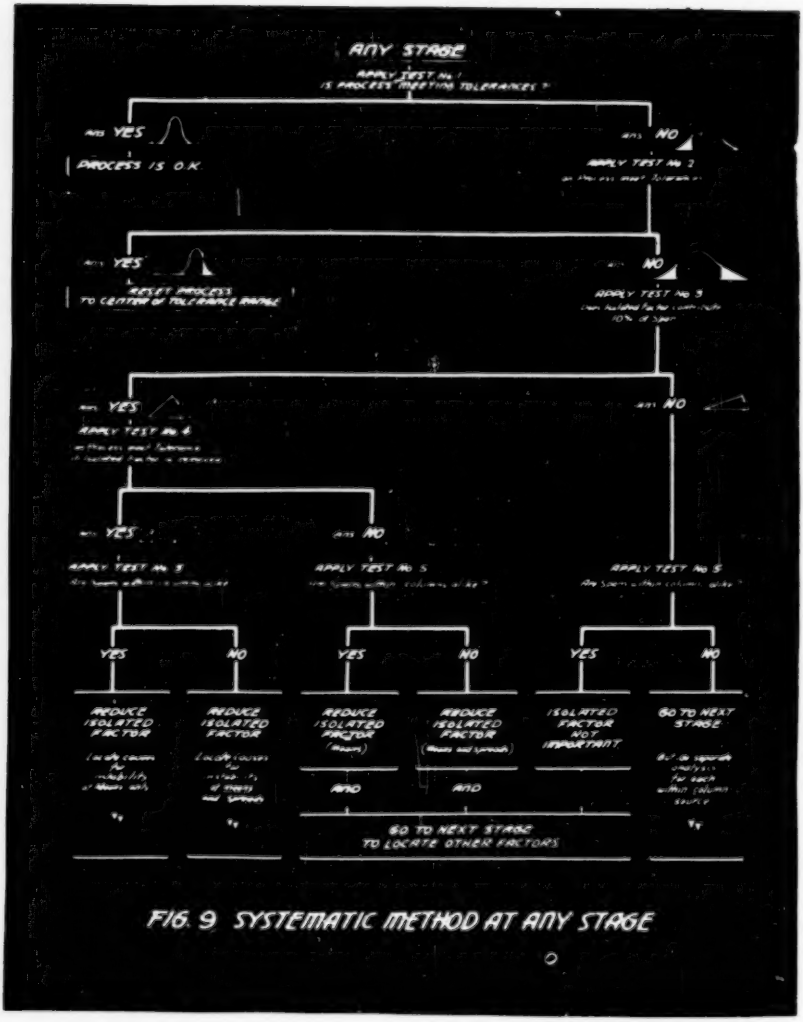


FIG. 9 SYSTEMATIC METHOD AT ANY STAGE

will be governed by this knowledge.

6. If Test #3 was positive, apply Test #4 to discover whether elimination of "lot-to-lot" variability would enable process to meet specifications.

   a. If so, the analysis points to a reduction in "lot-to-lot" factor as the necessary action. First, however, apply Test #5 to learn whether the within-lot variabilities may be considered as alike. If so, the investigation of "lot-to-lot" effects may be concentrated on causes of unstable lot averages. If not, causes for both unstable averages and dispersions must be unearthed. Follow the suggestions in step #7.

   b. If not, it means that "lot-to-lot" must be reduced AND the analysis must also proceed to Stage II to discover further reducible factors. As a practical matter, it may be well to proceed to Stage II, withholding judgement on the importance of "lot-to-lot" until the major factor is isolated. (Oftentimes, it will turn out that this major factor, deeper in the process, will have been responsible for the "lot-to-lot" indication at Stage I) Use Test #5 as in steps 5b and 6a before final action.

7. Suggestions for action toward reducing "lot-to-lot" variability:

   a. Because the lots were sampled over a period of time, the cause may be presumed to be associated with a variable that acted over the same time cycle. Some suspects are: time changes which affect the entire operation, such as very slow tool wear, comtamination, depletion of baths; differences between shifts of operators due to production or inspection habits; variation in lots of input material; changes in replaceable parts of equipment, such as dies, tools, molds, patterns, calender rolls.

   b. Consider how the lots are formed. Is the composition of a lot essentially determined at the operation being studied? (viz. a first operation converting rod or sheet into blanks or pieces; a machine being fed continuously with a sequence of earlier lots with no attention paid to these in the removal of completed lots). If so, it is highly probable that the "lot-to-lot" effect is caused at the operation itself.

   c. On the other hand, if the composition of the lot is essentially determined at an earlier operation (viz. "batch" processing of vacuum tubes, rectifiers, etc., in which lots are formed at an early operation and their exact identity is preserved through the remainder of the process), it is quite possible that the "lot-to-lot" effect is caused at a previous operation.

   d. A special case arises when the composition of individual lots is limited to single streams and there are two or more streams within the operation (viz. two parallel assembly lines each producing completed lots, or three presses performing the same operation each feeding full lots to the next operation). In such a case, proceed immediately to Stage II - the separate sampling of these streams - but return to analysis of the "lot-to-lot" factor if no significant stream effect is found.

e. Draw a control chart of the data at hand, compute limits; key in any identifying information originally supplied with the lots. This will provide a picture of the variation which, together with the supplementary information, may suggest one of the causes listed above in (a).

f. Continue the control chart as new lots are produced, gathering as much supplementary information as possible to study for correlation.

g. Make a control chart record of the input material and key it with the chart of the same lots after the operation. This may show that the "lot-to-lot" factor is essentially a reflection of trouble in an earlier operation. If so, move back to the earlier operation and subject it to the same kind of detailed analysis.

Stage II.

It is first necessary to consider the possibilities for stream effects which exist in the manufacturing process. See definition of "streams" for examples. The important idea is "parallel paths". Whenever the item has a "choice" of machines, equipment, fixtures, cavities, etc. through which it may go in being converted from input material to finished product, a potential stream effect is present. (Note that this concept applies even though product flows through only one stream at any one time, as in the case of a single punch-press having a standby replacement die. The fact that the second path exists means a possible stream effect. Unless this has already been included in the "lot-to-lot" factor of Stage I by the time interval over which the sampling was conducted, it should be considered as part of the "stream-to-stream" analysis.)

1. If Test #5 of Stage I was negative*, proceed as follows: Sample from each stream according to Procedure A and measure the relevant quality characteristic. Draw the sample in one of two ways:

   a. Accumulate a "sub-lot" from each stream, i.e. the number of items which that stream would normally contribute to a lot of outgoing product, of the size used in Stage I. Sample at random from within each of these "sub-lots".

   b. Draw the required number of items singly at random times from each stream. The total time interval over which the sample is collected should correspond to the time for the stream to produce a "sub-lot", as defined above.

2. Compute the Average and Total Span, obtain the UPL and LPL, compute the components of Total Span and draw the pictogram, as specified in Procedures B & C.

* If Test #5 of Stage I was positive, it means that process is subject to a source of variation whose presence may not be detected in the length of time for the stream to produce a sub-lot. Hence, lengthen the time interval for sampling in Step #1 or else repeat the entire Stage II procedure several times until the variability encountered in Stage I shows itself.

2. Compute the Average and Total Span, obtain the UPL and LPL, compute the components of Total Span and draw the pictogram, as specified in Procedures B & C.

3. Apply Test #1 to determine whether Total Span is meeting the specified tolerances.

   a. If so, it means that process is now satisfactory. However, since process must have been judged unsatisfactory to have reached Stage II, this appears contradictory. Apply Test #6 to learn whether the contradiction is real or the accident of sampling.

      (1) If Test #6 shows that the Total Span at Stage II is significantly smaller than the within-lot Span of Stage I, it means that the Stage II sample has not picked up a source of variation that was present at the time the Stage I Analysis was made. (e.g. perhaps a "stream-to-stream" effect was previously present but has now been corrected, consciously or unconsciously). If evidence of conscious correction can be found, the investigation should pursue the question of whether such correction is effective and adequate so that it may be specified for the future. If not, it is necessary to repeat the Stage II sampling, possibly extending the time interval of the collection of the sample, in an effort to capture the elusive factor.

      (2) If Test #6 shows that the Total Span at Stage II is significantly larger than the within-lot Span of Stage I, it indicates that the later sample has encountered a source of variation not present when the Stage I sample was taken. Proceed to step #4 and consider the Total Span of Stage II as being more representative of the true process spread than the Stage I figure. (There is an exception to this conclusion, however, when there are only 2 or 3 streams and a sizeable stream-to-stream effect is present. Under such conditions the Total Span estimate will be considerably inflated over the Stage I value. This is not serious since the stream effect will be clearly seen and isolated, but an exaggerated picture of its magnitude may be given. See step #8).

      (3) If Test #6 does not show a significant difference, proceed with Step 4 on the assumption that the Total Span obtained may well be an underestimate and that it is still necessary to isolate a source of variability even though the process is <u>apparently</u> meeting tolerances.

   b. If not, proceed to step #4.

4. Apply Test #2 to determine whether Total Span is capable of meeting the specified tolerance range.

   a. If so, the required action is apparently to reset the mean of the process at or near the midpoint of the tolerance range and consider the analysis as complete. However, since process must have been judged incapable to have reached Stage II, this appears contradictory. Follow the procedure suggested under Step #3a above.

b. If not, proceed to step #5.

5. Apply Test #3 to determine whether the "stream-to-stream" factor contributes 10% or more to the Total Span.

   a. If so, proceed to step #6.

   b. If not, proceed to Stage III. However, it is first necessary to apply Test #5 to learn whether the variabilities within each stream may be considered to be alike, since the procedure at Stage III will be governed by this knowledge.

6. If Test #3 was positive, apply Test #4 to discover whether elimination of "stream-to-stream" variatibility would enable process to meet specifications.

   a. If so, the analysis points to reduction in the "stream-to-stream" factor as the necessary action. First, however, apply Test #5 to learn whether the within-stream variabilities may be considered as alike.

      (1) If so, the investigation of "stream-to-stream" effects may be concentrated on causes of unstable stream <u>averages</u>.

      (2) If not, investigation must consider causes for both unstable <u>averages</u> and <u>dispersions</u>.

   b. If not, it means that "stream-to-stream" must be reduced AND the analysis must also proceed to Stage III. Follow suggestion given in step #6b of Stage I.

7. Suggestions for action toward reducing "stream-to-stream" variability:

   a. The "stream-to-stream" effect must necessarily be associated with some physical difference between the paths taken by the items of product. If the analysis has disclosed some "good" and some "bad" streams, advantage can often be taken of this knowledge by making independent comparisons of physical measurements of the different equipments involved. What physical characteristic is possessed by the equipment producing at the "good" level that is different from the "bad" level equipment?

   b. If the observed effect is that of different averages of the different streams, (i.e. if Test #5 was positive) the most obvious source is the particular feature of the equipment that determines the "setting" of the process. Examples: tool settings, ram settings, dimensions of dies or cavities, temperature settings, bath concentration levels, speed settings, systematic instruments calibration errors, differences in operator skills in attaining the target setting.

   c. Collect additional data on the stream effect to discover whether the differences between streams remain constant or fluctuate over a period of time. The former points to a permanent dimensional difference -- like a die size -- while the latter suggests that the real problem is control of time effects within the streams. In the latter case, proceed to a Stage III analysis.

d. Investigate manner in which input material is fed into streams. Does each stream get a series of lots from a particular stream of an earlier operation? If so, the "stream-to-stream" effect at hand may in reality reflect a stream effect in an earlier operation. Move the analysis back to the earlier operation.

8. In step #3a-2, it was pointed out that an inflated estimate of Total Span may result when there are only 2 or 3 streams and a large stream-to-stream effect is present. This is because of the tacit assumption of the analysis that the streams sampled are really randomly selected from an infinite number of possible streams. Sometimes this model actually fits the situation, as when the setting of each stream is independently adjustable and is likely to change frequently (i.e. when each stream is produced by a different tool setting, each of which is separately subject to tool wear, slippage, operator adjustment, etc.) In such a case, the "inflated" estimate is really a true estimate, predicting the variability that will result in the future if no control is exerted on the settings of the stream. However, when the stream settings are essentially fixed and non-adjustable (i.e. dimensions of molding cavities or punch press dies), such a prediction is usually not of interest; rather it is desired to estimate the Total Span, given the existing differences in stream settings. In such a case, it is suggested that all of the Stage II data be put into a frequency distribution and the Total Span estimated visually by the spread of the distribution.

Stage III

At this stage, the object is to isolate short-term time effects within a stream. If Test #5 of Stage II has shown no difference in variability within the streams, Stage III may be applied to only one of the streams. If Test #5 has shown such differences, Stage III should be applied separately to each of the streams, or at least to those which showed the larger variabilities in Stage II.

1. Sample successive pieces from the stream at 10 different times, as specified in Procedure A. The total time interval over which the sampling is performed should correspond to the time required for production of a sub-lot, as defined in Stage II, step #1.

2.
3.
4. ——— Proceed as in Stage II, except substitute "time-to-time" for "stream-to-stream", and "within-time" for "within-stream". (The remarks concerning small k's under step #3a-2 will not apply to this stage since k = 10).
5.
6.

7. Suggestions for action toward reducing "time-to-time" variability:

   a. The time effects being searched out here are generally of shorter duration than those sought in the Stage I analysis and are focussed on a much smaller segment of the process. Examples: rapid tool wear (taking place within a sub-lot), tool slippage, process adjustment by the operator, bar-to-bar variation within

a lot of raw material, changes in machine settings due to warm-ups, interruptions, rest periods, etc. However, the techniques suggested under Stage I, steps 7e, 7f, 7g are applicable here, by simply reducing the time scale.

Stage IV

This stage, while not so universal in its application to all processes as the preceding stages, nevertheless is applicable in a variety of mechanical operations. Generally, it is applicable in any situation where it is possible to obtain readings at different positions on a single piece. When these readings differ by more than simply the error of measurement, there is said to be a "piece-positional" effect. It is, in other words, a systematic difference or stratification of readings around the periphery along the edge or across the surface of a piece. (Readings may be of dimensions or of a physical characteristic, like hardness or density.) For example, diameters are usually out-of-round, cylindrical surfaces may have taper along their lengths, top and bottom surfaces of a flat piece are never quite parallel, holes have bell-mouths or back-tapers, etc.

These conditions are usually traceable to faults in the size, shape or set-up of the tooling and it is therefore the purpose of Stage IV to split off this "within-piece" variability and measure its magnitude compared to the remaining unresolved factor, the "piece-to-piece" variability.

Stage IV is to be applied to any one set of consecutive pieces from any one stream, provided Test #5 has been positive in both Stages II and III. If Test #5 has been negative in either stage, Stage IV should be repeated for each stream or for each stream and/or each time period that has given suspicious variability.

1. Sample n successive pieces from the stream at one particular time, as specified in Procedure A. A portion of the sample already taken for Stage III is suitable if it has not been destructively tested. Measure each piece at k different positions. In deciding the number and location of the readings, the type of operation must be considered. When the piece can be oriented with respect to some reference point, it is desirable to designate definite locations and to measure at the same locations on each piece. When the piece cannot be oriented, (as with most diameters), take an infinite number of readings (by rotating the diameter under a dial indicator, for example) and record only the maximum and minimum readings.

2.
3.
4. ——————— Proceed as in Stage II, substituting "piece-positional" for "stream-to-stream" and "piece-to-piece" for "within-stream".
5.
6.

7. Suggestions for action toward reducing "piece-positional" variability:

   a. The tooling should be subjected to close scrutiny by inspection against the tool specification. It may be worn, bent or otherwise damaged.

b. Consideration should be given to the method of mounting or holding the tool. Can the mounting be strengthened? Can the tool post be moved nearer to the chuck to get a steadier action? Can the holder be adjusted to give a better alignment between the tool and work? Can a gage be devised to check work-tool alignment before producing parts?

c. In the case of "piece-positional" variability on flat molded, calendered, pressed or stamped work, the problem generally stems from alignment faults between mold halves, calender rolls, punch and die, etc. Sometimes these faults can be corrected by precise measurement of the alignment of the tools themselves, but more often the extent is too small to be readily picked up in this way. However, the measurements already made on the piece-parts are extremely valuable in seeking out the alignment correction. By arranging the readings in a pictorial pattern, known as a "Position-Dimension Diagram",(7) it is often possible to make very accurate diagnoses. These diagrams also offer a convenient check for each new set-up to assure minimum "within-piece" variability.

d. When a "piece-positional" effect has been found in the particular stream sampled, all other streams should be subjected to Stage IV analysis and appropriate corrective measures taken to reduce this type of variability throughout the process.

Stage V.

In most mechanical processes, the precision of measurement involved in the gage or measuring device is small enough to be negligible from a practical standpoint. However, this cannot be blithely assumed, particularly if the process has resisted all the efforts of the previous four stages to yield a reducible source of variation. For this reason, Stage V is intended to segregate the true product variability from the error associated with the precision of measurement. (Note that systematic errors in calibrating or reading the instrument are not included -- only those random differences among repeat readings commonly known as "observational error"). It can obviously be applied only in the case of a non-destructive measurement.

Stage V is to be applied to any one set of consecutive pieces taken from any one stream, assuming that Test #5 has been positive in Stages II, II and IV. Fortunately, however, even if Test #5 has been negative, it will not affect the estimate of Span of the "error of measurement"; it will affect only the estimate of "inherent" Span. Thus, if any Test #5 has been negative, use the Special Directions of step #7, below, for making Tests #3 and #4.

1. Sample k consecutive pieces from any one stream at a particular time, as specified in Procedure A. A portion of the sample used for Stage III is suitable. Select one particular position and take n repeat measurements at the same position on each piece. To avoid bias due to memorizing of readings, it is well to take "first" readings on all of the pieces, then shuffle the order of the pieces and take all the "second" readings, etc. (Retain the separate identification of each piece, however.)

2. Compute the Average and Total Span, obtain the UPL and LPL, compute the components of Total Span and draw the pictogram, as specified in Procedures B and C. However, it must be pointed out that Stage V deviates from the rule that the among-column effect is the isolated factor and the within-column effect is the unresolved factor. Here the k pieces measure the "inherent" or unresolved factor, while the n repeat readings measure the "error of measurement" or isolated factor. To avoid confusion, it is suggested that the $3\ \sigma_{if}$ and $3\ \sigma_{un}$ values be computed on the worksheet as with the previous stages, but that care be taken to note the special mathematical treatment for this stage, as detailed below. In drawing the pictogram, reverse the usual positions of $3\ \sigma_{if}$ and $3\ \sigma_{un}$ to correspond with the labelling shown in Fig. 1.

3. Apply Test #1 to determine whether Total Span is meeting the tolerances.

   a. If so, it means that process is now satisfactory. However, since process must have been judged unsatisfactory to have reached Stage V, this appears contradictory. Apply Test #6 to learn whether the contradiction is real or the accident of sampling.

      (1) If Test #6 shows that the Total Span at Stage V is significantly larger or smaller than the piece-to-piece Span of Stage IV, it probably means that the Stage V sample is not truly representative of the piece-to-piece variability. Proceed with the Special Direction of step #7, below.

      (2) If Test #6 does not show a significant difference, proceed to step #4 on the assumption that the Total Span of Stage V is an underestimate and it is still necessary to evaluate the error of measurement even though the process <u>appears</u> to be meeting the tolerances. In some cases, it may be wise to use Special Directions in step #7, below.

   b. If not, proceed to step #4.

4. Apply Test #2 to determine whether Total Span is capable of meeting the tolerances.

   a. If so, the required action is apparently to reset the mean of the process at or near the midpoint of the tolerance range and consider the analysis as complete. However, since process must have been judged incapable to have reached Stage V, this appears contradictory. Follow the procedure suggested under step #3a, above.

   b. If not, proceed to step #5.

5. Apply Test #3 to determine whether the "error of measurement" factor contributes 10% or more to the Total Span. Modify the ratio for Test #3 by using $\frac{3\ \sigma_{sp} - 3\ \sigma_{if}^{*}}{3\ \sigma_{sp}}$ instead of $\frac{3\ \sigma_{sp} - 3\ \sigma_{un}}{3\ \sigma_{sp}}$.

* Worksheet nomenclature; actually "inherent" Span.

a. If so, proceed to step #6.

b. If not, error of measurement is not a major factor in the Total Span. See "After Stage V" for further action.

6. If Test #3 was positive, apply Test #4 to discover whether elimination of the "error of measurement" would enable process to meet specifications. Use 6 σ_{If}* instead of 6 σ_{un} to compare "inherent" span with tolerances.

a. If so, it will be necessary to reduce the error of measurement. See step #8 for suggestions.

b. If not, it may still be desirable to reduce error of measurement. Also, the wisdom of the tolerances may well be questioned. See "After Stage V" for further action.

7. Special Directions. If there has been any suggestion at step #3 that the sample does not fairly represent the "piece-to-piece" Span of Stage IV, it is likely that the "inherent" Span estimate is faulty, since it depends on a relatively few pieces. It is therefore suggested that, in carrying out steps #5 and #6, a revised estimate of "inherent" variability -- based on the more reliable Stage IV data -- be used. Proceed as follows:

a. Let (3 σ_{un})IV represent the unresolved or "piece-to-piece" Span obtained at Stage IV (or Stage III if IV was not applicable).

Let (3 σ_{un}**)V represent the "error of measurement" Span obtained at Stage V.

b. Compute a new estimate of (3 σ_{If}*)V, the "inherent" Span at Stage V, from the formula:

$$(3\ \sigma_{If}^{*})V = (3\ \sigma_{un})^{2}IV - (3\ \sigma_{un}^{**})^{2}V$$

c. Construct a revised pictogram for Stage V, using:
$\bar{\bar{X}}$ from Stage Iv
(3 σ_{un})IV as the Total Span
(3 σ_{If}*)V (Revised) as the "inherent" Span
(3 σ_{un}**)V as the "error of measurement" Span.

d. Now perform step #5 of the Stage V procedure.
The ratio for use in Test #3 will be:

$$\frac{(3\ \sigma_{un})IV - (3\ \sigma_{If}^{*})V \text{ (Revised)}}{(3\ \sigma_{un})IV}$$

e. If step #6 is used, then Test #4 is modified by comparing 2 x (3 σ_{If}*)V (Revised) with UTL - LTL.

8. Suggestions for improving error of measurement:

a. The most obvious action is to obtain a more precise piece of measuring equipment. This may require the advice of an expert or

* Worksheet nomenclature; actually "inherent" Span.
** Worksheet nomenclature; actually "error of measurement" Span.

reference to a good text on modern gaging practices, such as Kennedy[8]. Representative of instrument manufacturers may be helpful.

b. The large error may be due to the ineptness of the observer rather than the inherent precision of the instrument. Repeat Stage V with several observers, preferably with the same piece-parts, and note whether $3\sigma_{un}$ is appreciably different for the different observers.

c. When the error of measurement is really sizeable, it will pay to divide the measuring operation into elements and determine the separate contribution of each element to the total. Such elements might be, for example; error of positioning workpiece on supports, error of positioning dial indicator on workpiece, error of zero-setting the indicator, error of reading the indicator, etc. With a little ingenuity in designing such experiments, they can readily be adapted to SPAN PLAN analysis.

d. If all else fails, error of measurement can always be reduced by the simple -- though costly -- expedient of requiring several repeat readings to be taken of the same measurement and using the average of these. To estimate the number of such repeat readings necessary for a given reduction in error, note the following table:

To reduce 3σ error to:	Required repeat readings
1/2 of present error	4
1/4 of present error	16
1/10 of present error	100

After Stage V

The unresolved factor which remains after completion of Stage V must, in the present method, be regarded as "inherent" or "natural". Broadly, this means that a fundamental change in the process is required if it is still unable to meet the tolerances. However, even at this point, there is a challange to the ingenuity of the investigator to invent simple changes in the equipment that may be fundamental enough in their influence on the reproducibility of pieceparts to provide substantial reduction in "inherent" variability. In the authors' experience, there have occurred numerous examples where such things as the addition of a positive stop to a springloaded tool or the provision of tighter work-clamping device have successfully reduced "inherent" variability by 50%.

As the "irreducible minimum" is approached, of course, the question of whether the tolerances are rational should be raised. Reconsideration of tolerances by the design engineering function, based on the factual information on capability that these procedures have yielded, is a legitimate means for solving the defect prevention problem.

III. DETAILED PROCEDURES FOR EVERY STAGE

Procedure A

Collecting the Experimental Data

Object: To specify sampling point, method of sampling, number of samples and size; also to obtain measures of Total Span and the two factors which comprise it at any stage.

Method:

1. Refer to Table A-1 and, entering it at the appropriate stage, read off the sampling point and method of sampling.

2. Determine k, the number of samples to be taken. This is the number of different lots, streams, times or piece-positions (depending on the stage) which are to be sampled. In general, the choice of k will be dictated by the circumstances of the process, such as the number of streams which actually exist or the number of piece-positions which are usually measured. In Stages I, III, and V, however, where k denotes the number of lots, times, and pieces, respectively, the choice is an arbitrary one. In such cases, k should be chosen as 10. It is best to avoid small k's whenever possible. Whenever the "natural" number dictated by the process exceeds 10, it is suggested that the number be reduced to 10 by random selection from the actual number available.

3. Refer to Table A-2 for the value of n which is appropriate for the k which has been chosen. This tells how many items or pieces are to be selected within each of the k lots, streams, etc.

4. Perform the sampling and measurement and enter the results in an array of k columns and n rows.

Procedure B

Computing Average Total Span and Components of Total Span

Object: To utilize the data collected to obtain estimates of the centering and Total Span of the process and the magnitude of the two components of Total Span at any stage.

Method:

1. Compute the sum of each of the k columns.

2. Add these sums to obtain the grand sum. (Σx).

3. Divide the grand sum by kn to obtain the average. ($\bar{\bar{X}}$)

4. Compute the range within each column. (R_{un})
   (This range yields the estimate of the unresolved factor; hence the subscript "un").

5. Add these k ranges together. (ΣR_{un}).

Table A-1

Sampling Specifications for each Stage

Stage	Sampling Point	Method of Sampling	Measurement of Pieces	k	n
I	From completed lots as they emerge from operation or enter the following operation.	Random within the lot	At random position on each piece.	No. of lots	No. of pieces from a lot
II	From individual streams before they are merged to from lots.	Random from accumulated product of a stream, or at random times during production.	At random position on each piece.	No. of streams.	No. of pieces from a stream.
III	From individual streams before they are merged to form lots.	In sequence within the streams at periodic time intervals.	At random position on each piece.	No. of times samples are drawn.	No. of pieces at a time.
IV	From individual streams before they are merged to form lots.	In sequence within the stream at periodic time intervals.	At specified positions on each piece. (Except out-of-roundness, eccentricity, etc. where "max" and "min" are recorded)	No. of positions measured on each piece.	No. of pieces checked at each position.
V	From individual streams before they are merged to form lots.	In sequence within the stream at a particular time.	At one specified position on each piece. Take n repeat readings.	No. of pieces measured.	No. of repeat readings on each piece.

Table A-2

No. of items required within each sample (n)	No. of independent samples to be taken (k)								
	2	3	4	5	6	7	8	9	10
	10	10	9	8	7	7	6	6	6

6. Obtain appropriate value of C_{un} from Table B-1 and divide it into the sum obtained in step #2 above to obtain the Half-Span of the within-column (unresolved) factor ($3\sigma_{un}$).

7. Now compute the range of the column sums. (R_{if}). These sums were obtained in step #1.
(This range yields the estimate of the isolated factor; hence the subscript "if").

8. Obtain the Half-Span of the among-column (isolated) factor ($3\sigma_{if}$) by means of the following formula:

$$3\sigma_{if} = \frac{1}{\sqrt{n}} \sqrt{\left(\frac{R_{if}}{C_{if}}\right)^2 - \left(3\sigma_{un}\right)^2}$$

where C_{if} is obtained from Table B-2. If the number under the square root is negative, $3\sigma_{if}$ is zero.

9. Compute the Half-Span of the Total Span ($3\sigma_{sp}$) by means of the following formula:

$$3\sigma_{sp} = \sqrt{\left(3\sigma_{if}\right)^2 + \left(3\sigma_{un}\right)^2}$$

10. Add this Half-Span ($3\sigma_{sp}$) to $\bar{\bar{X}}$ (Step #3) to obtain the Upper Process Limit (UPL).

11. Subtract this Half-Span ($3\sigma_{sp}$) to $\bar{\bar{X}}$ to obtain the Lower Process Limit (LPL).

Table B-1

k-n Combination Being Used

Conversion factor for "un" Half-Span.

k n	2 10	3 10	4 9	5 8	6 7	7 7	8 6	9 6	10 6
C_{un}	2.08	3.12	4.00	4.80	5.46	6.37	6.80	7.65	8.50

Table B-2

k-n Combination Being Used

Conversion factor for "if" Half-Span.

k n	2 10	3 10	4 9	5 8	6 7	7 7	8 6	9 6	10 6
C_{if}	1.49	2.01	2.24	2.34	2.36	2.50	2.42	2.52	2.60

Procedure C

Constructing the Pictogram

Object: To construct a Pictogram showing the relationship between the Total Span at any stage and Spans of the isolated and unresolved components of the Total Span. Also, to indicate the location of the average and the process limits with respect to the required tolerances.

Method:

1. Lay out a scale in the units of measurements which were used to measure the samples. The scale must be long enough to cover the distance between the upper and lower tolerances or between the upper and lower process limits, whichever is larger.

2. Locate the average ($\bar{\bar{X}}$) on this scale.

3. Using a compass, with $\bar{\bar{X}}$ as a center, lay off a distance on the scale equal to $+3\ \sigma_{sp}$ and another distance equal to $-3\ \sigma_{sp}$. These points are the UPL and LPL respectively.

4. With $\bar{\bar{X}}$ as a center and a radius equal to $3\ \sigma_{un}$, inscribe two arcs above and near the UPL and LPL.

5. With the UPL as a center and a radius equal to $3\ \sigma_{if}$, inscribe an arc across the previous arc. Connect this intersection with straight lines to the UPL and to $\bar{\bar{X}}$.

6. With the LPL as a center and a radius equal to $3\ \sigma_{if}$, inscribe an arc across the other previous arc. Connect this intersection with straight lines to the LPL and to $\bar{\bar{X}}$.

7. Erect vertical lines at the tolerance values.

8. Label the lines as shown below in Fig. 10, using the actual names of the isolated and unresolved factors.

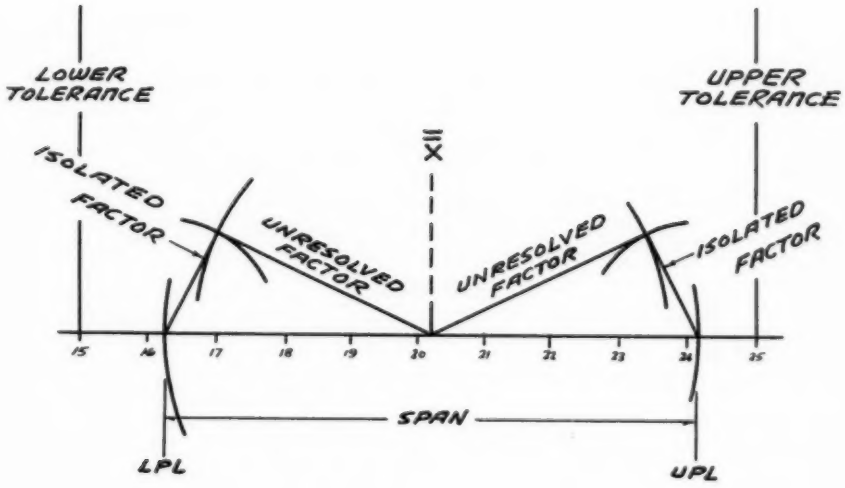


Fig. #10

IV. TESTS FOR EVERY STAGE

Test #1

Is Process Meeting Tolerances at this Stage?

1. Compare location of UPL to Upper Tolerance and location of LPL to Lower Tolerance.
2. Answer is YES if both process limits are within tolerances.
3. Answer is NO if either process limit is outside of tolerances.

Test #2

Can Process Meet Tolerances at this Stage?

1. Obtain the Total Span by doubling the Half-Span found in step #9 of Procedure B.
2. Subtract the Lower Tolerance from the Upper Tolerance to obtain the Tolerance Range.
3. Compare the Total Span with the Tolerance Range.
4. Answer is YES if Total Span is equal to or less than Tolerance Range.
5. Answer is NO if Total Span is greater than Tolerance Range.

Test #3

Does Isolated Factor Contribute 10% or more to Total Span?

1. Compute the ratio: $$\frac{3\sigma_{sp} - 3\sigma_{un}}{3\sigma_{sp}}$$

   Use values of $3\sigma_{un}$ and $3\sigma_{sp}$ from steps #6 and #9 of Procedure B.
2. Answer is YES if ratio is .10 or more.
3. Answer is NO if ratio is less than .10.

Test #4

Can Process meet Tolerances if Isolated Factor is Removed?

(To be used only if Test #3 is positive)

1. Obtain the Span of the unresolved factor by doubling the within-column Half-Span found in step #6 of Procedure B.
2. Subtract the Lower Tolerance from the Upper Tolerance to obtain the Tolerance Range.
3. Compare this Span to the Tolerance Range.

4. Answer is YES if Span is equal to or less than Tolerance Range.

5. Answer is NO if Span is greater than Tolerance Range.

Test #5

Are Spans Within-Columns Alike?

1. Divide the range of the R_{un} values (ρR_{un}) by 3 σ_{un}.

2. Compare this quotient with the number Q for the appropriate k-n combination being used. (See Table of Q below).

3. Answer is YES if $\frac{\rho R_{un}}{3\,\sigma_{un}}$ is less than Q.

4. Answer is NO if $\frac{\rho R_{un}}{3\,\sigma_{un}}$ is equal to or greater than Q.

k	2	3	4	5	6	7	8	9	10
n	10	10	9	8	7	7	6	6	6
Q	.57	.78	.91	1.01	1.09	1.15	1.22	1.24	1.27

Test #6

Does Total Span at this Stage Differ from Unresolved Span at Preceding Stage?

1. Divide 3 σ_{sp} of the present stage by 3 σ_{un} of the preceding stage. Call this ratio: SP/UN.

2. Refer to the value of $\frac{3\,\sigma_{sp} - 3\,\sigma_{un}}{3\,\sigma_{sp}}$ that was calculated in step #1 of Test #3..

3. Enter the table of Upper Limits below and find the Upper limit for SP/UN corresponding to the value of k at the preceding stage and to the nearest value of $\frac{3\,\sigma_{sp} - 3\,\sigma_{un}}{3\,\sigma_{sp}}$ at the present stage.

4. If the value of SP/UN equals or exceeds its upper limit, then the Total Span at this Stage is significantly GREATER than the Unresolved Span at the preceding Stage.

5. Obtain the lower limit for SP/UN from the Lower Limit table in the same manner as in step #3.

6. If the value of SP/UN is equal to or less than its lower limit, then the Total Span at this Stage is significantly LESS than the Unresolved Span at the preceding stage.

7. If the value of SP/UN lies within the upper and lower limits, then the Total Span at this Stage does not differ from the Unresolved Span at the preceding stage.

UPPER LIMITS FOR SP/UN						
Value of $\frac{3\,\sigma_{\bar{sp}} - 3\,\sigma_{un}}{3\,\sigma_{\bar{sp}}}$ at present stage	Value of k at preceding stage					
	2	3	4	5-6	7-8	9-10
0	1.59	1.54	1.52	1.51	1.49	1.48
.10	1.66	1.58	1.54	1.52	1.50	1.49
.20	1.73	1.66	1.60	1.57	1.54	1.52
.30	1.94	1.75	1.66	1.62	1.58	1.55
.40	2.05	1.83	1.73	1.67	1.62	1.59
.50	2.15	1.90	1.79	1.73	1.65	1.62
.60	2.22	1.95	1.85	1.77	1.68	1.64
.70	2.29	2.00	1.90	1.80	1.72	1.66
1.00	2.40	2.09	1.94	1.83	1.75	1.70

LOWER LIMITS FOR SP/UN						
Value of $\frac{3\,\sigma_{\bar{sp}} - 3\,\sigma_{un}}{3\,\sigma_{\bar{sp}}}$ at present stage	Value of k at preceding stage					
	2	3	4	5-6	7-8	9-10
0	.61	.66	.68	.69	.70	.71
.10	.52	.62	.65	.67	.69	.70
.20	.39	.53	.59	.62	.65	.67
.30	.28	.44	.52	.57	.61	.64
.40	.16	.36	.46	.52	.57	.61
.50	.05	.28	.40	.46	.53	.57
.60	.04	.25	.35	.44	.50	.54
.70	.04	.22	.29	.40	.47	.52
1.00	.03	.16	.27	.37	.43	.51

References and Notes

1. Seder, L. A., "Diagnosis with Diagrams"; Industrial Quality Control, Vol. VI, No. 4 (January, 1950)

2. Ibid

3. The word "Span" (or more accurately, "Total Span") is used generally to designate the total variability at any of the five stages. This is done to avoid using the word "capability" in several different senses.

4. David, H. A., "Further Applications of the Range to the Analysis of Variance"; Biometrika 38, 393-409 (1951)

5. Hartley, H. O., "The Use of Range in Analysis of Variance"; Biometrika 37, 271-280 (1950)

6. Seder, L. A., "Diagnosis with Diagrams", loc. cit.

7. Ibid

8. Kennedy, C. W., "Inspection and Gaging"; Industrial Press N.Y. (1951)

STATISTICAL QUALITY CONTROL IN THE PLASTICS INDUSTRY

Roger C. Boyd
Bakelite Company
A Division of Union Carbide and Carbon Corporation

During the past few years, STATISTICAL QUALITY CONTROL has won for itself a commanding name in the plastics industry as it has in most industries of today. The subject of STATISTICAL QUALITY CONTROL relative to the Plastic Industry presents a number of difficulties, since this industry is comprised of both science and art with secrecy surrounding portions of each. This is evident in the case of individual manufacturers, as well as selected groups of materials. It should be pointed out that statistical methods per se are not the universal answer which many uninitiated people believe. Rather, the successful solution of various manufacturing problems and the resultant control of variables not before recognized depends on common sense, sound engineering knowledge of the product and process involved, and development of good testing methods. These three factors can be aided considerably by the intelligent application of statistics.

Although Bakelite Company manufactures a variety of plastic materials including phenolics, vinyls, polyethylene, polystyrene, and fluorothene, we will be directly concerned with operations which convert or compound vinyl and polyethylene resins into sheet or granular materials. Familiar products fabricated from vinyl resin compounds are shower curtains, raincoats, furniture upholstery, and electrical insulation for house wiring. The popular "squeeze bottle", insulation for television antenna lead-in wires, refrigerator bowls, and thin film for a variety of packaging uses are examples of applications involving polyethylene resins and compounds. Thus, the ultimate market for our basic materials is a combination of consumer and industrial users, and there is considerable difference between the two, as many of you are probably aware. While both markets may be concerned with the same properties relative to any given product, the degree of importance attached to any one property may differ appreciably from one to the other.

As is the case with some consumer products and processes, far too many of the properties of the finished article must be judged, estimated, or arbitrarily rated by visual means. Granted, there are statistical methods available by means of which such situations may be treated, but these are not too helpful if the original parameters or testing methods themselves are weak. The reduction of problems, such as this, to a quantitative measurement is the object of a continuing search within this and many other industries.

An example of this is color control. Flexible vinyl film is used almost exclusively in the consumer trade and is furnished in an infinite variety of colors, surfaces, and patterns. Strict control of color is mandatory in this case since color-matching with other materials used in the home is continually involved. On the other hand, more latitude in color specifications is allowable in the case of wire and cable materials furnished to the industrial market. In this case, color is used almost solely for coding purposes in multiple conductor

constructions. The problem reduces itself to one of being able to furnish a small number of colors where blue, white, red or orange can be easily differentiated from one another.

Fortunately, science has recently been recognized in the field of color control with the development and application of the recording spectrophotometer. Using this instrument, it has been possible to establish a truly quantitative method of evaluating the color of a given material, thus opening the door for finite statistical control methods.

In a materials manufacturing industry such as ours, the problem of contamination, both functional and visual, is continually present. Normally, functional contamination such as the presence of small amounts of a conductive component in an electrical insulation material is controllable by quantitative electrical measurements, and simple variables control charts can be used. On the other hand, visual contaminants in a consumer material, which affect only its appearance, present a widely different problem. Here, one must either classify contaminants by type and concentration or deal with the "mass impression" of a given area of sample. Both approaches are commonly used depending upon the ultimate application of the material involved. In one case that comes to mind, the problem resolved itself to the simple task of counting the number of foreign particles in a given area of sample and using a variables control chart to catalogue process trends.

Of particular interest to any manufacturer, once the measurement of a defect has been achieved, is the problem of either completely eliminating it, or at least minimizing it, to the extent that the product is commercially salable. This problem is often one which can have far reaching consequences in an industry such as ours. One such case involved the sudden appearance of many black specks in a vinyl plastic sheet material. These specks, which had been determined to be magnetic particles, were in such volume that the appearance of the sheet was seriously impaired. Accordingly, the process was shut down, and all processing equipment was scrupulously examined to find the source of the contamination. This search was fruitless. Hence, all the raw materials (5 in number) were reinspected visually. This procedure yielded nothing. Yet, when a simple hand magnet was passed through each of them, one showed minor evidence of magnetic metal. More of this single material was passed through a magnetic separator, and the actual amount of tramp metal contained in it was determined to be on the order of ten parts per million. With the knowledge of the formulation of the finished product, this was back calculated, and it was thus determined that our source of contamination had been isolated. The vendor of the particular raw material was contacted and, after investigating his own material on hand, shut down his process to determine the origin of the particles. This was shortly determined to be due to faulty metal used in the makeup of the hammers in a conventional hammer mill. The result of all this was the following:

1. Our raw material inspection for this and similar materials was revised to include controls which would quantitatively measure the contaminant, and a maximum specification was established.

2. The particular vendor installed a similar control on his finished product and simultaneously reinforced his specification for the type and hardness of metal to be used in his grinding equipment.

3. The vendor installed magnetic separators in his manufacturing operation.

4. All other vendors of similar materials were contacted concerning the new specification for contamination, and they in turn established controls accordingly.

Of course, statistical calculations were involved in the latter stages of this problem in setting the specification and establishing the controls in both the vendor's plant and in our own. The largest contribution to the solution of the problem was common sense and general technical knowledge.

Probably of greater interest to you is the following problem which clearly could not have been solved without the use of statistics.

Difficulty was being encountered with a particular sheeting process wherein the processability or "formability" of batches of the crude mix varied. Accordingly, all the known variables having to do with the process were detailed. These amounted to approximately thirty. With the aid of background information and general engineering knowledge, three of these variables were estimated to be of the greatest importance. A laboratory experiment utilizing a Latin Square design was performed using small equipment. These three variables were set up in the experiment, each at two different levels. The property being measured was termed "Processability". By simply averaging the various components of the Latin Square, it was seen that two of the variables were of importance, whereas one was not. In order to confirm this conclusion, the levels of the one variable were changed to more closely approximate production conditions and the experiment was expanded to include all combinations of the three variables at two different levels. Thus, a complete factorial experiment was performed. It was finally found that all three variables were significant and that no important interactions existed. On the basis of this small amount of laboratory work, a factorial experiment involving full scale production equipment was planned and run. It is significant to note that in all cases the laboratory experiment was confirmed and the resulting process and raw material controls materially reduced processing variabilities. The contribution of statistical methods was of great importance in this case, since it enabled us to evaluate a number of variables simultaneously with a relatively small amount of experimentation. More important yet was the fact that this well-qualified laboratory work set up the boundaries for a production scale experiment wherein no material loss was sustained. Far too often, unqualified experimental work in the laboratory results in large material losses when an experiment graduates to the production scale.

It is difficult to compare quality control as practiced in the plastics industry to quality control as it is practiced in other older manufacturing enterprises. The plastics materials that we have discussed here have been in commercial production for only fifteen years at the most, and basic understanding of them is lacking as regards various properties both known and unknown. Thus, a fair portion of our quality control work is of a developmental nature in attempting to achieve a better understanding of our raw materials, processes, and finished products. It is widely thought that since the plastics industry is young, and is staffed largely by engineers and competent

technical help, that all is known about the materials being manufactured. This is definitely not the case. At times we, who live with it daily, wonder if we even have begun to understand these new materials. In such a situation, the use of the simple Latin Square and also complete factorial designs yield great dividends when applied to specific problems in process or materials analysis. Variables charts are used particularly in regard to "controlling" or primary properties. Statistical sampling is not widely used because batch processes are normally the case. This type of sampling is used in the case of some raw materials which are produced by a continuous process.

In conclusion, the plastics industry is in a rapidly expanding and growing stage. This applies to production of existing materials, development of new materials, and basic knowledge applicable to both. Due to this expansion, statistical quality control methods have a very definite place in the scheme. However, it is repeated that these methods must be applied intelligently, with due consideration being given to three very important items; namely, common sense, sound engineering knowledge of the product and process, and good testing methods.

SOME APPLICATIONS OF STATISTICAL QUALITY CONTROL IN THE PAPER BOXBOARD AND PRINTING INDUSTRIES

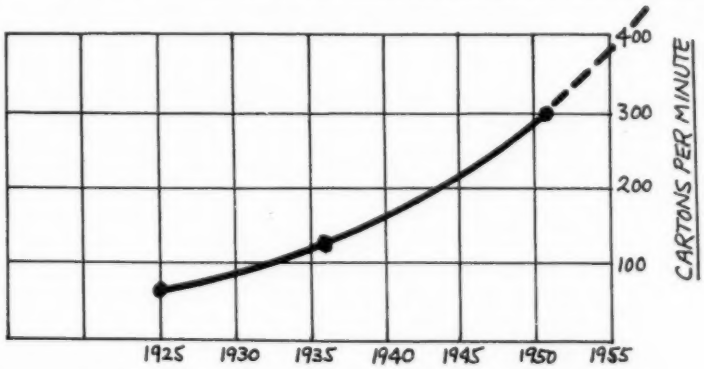
Brainard E. Sooy
The Gardner Board and Carton Co.

PREFACE

The following brief presentation of a few applications of Statistical Quality Control to the Folding Boxboard and Carton Printing Industries has been purposely kept as non-technical as possible. I well remember my own confusion when I started to study the subject, and it is hoped that this simplified pictorial presentation will make at least a part of the overall picture easier to understand.

While attending the Milwaukee Convention of the A.S.Q.C. three years ago I overheard one man say, "The machine tool people have done a wonderful job on Statistical Quality Control, but where can I find something on it that works in my business?" The answer is that there is probably nothing - one must master the fundamentals and then work out the applications to fit his own peculiar needs. The most important point is that the fundamentals will work in <u>any</u> business, and there is a real satisfaction in harnessing them to one's own needs.

WHY DO WE NEED QUALITY CONTROL?



STOKES & SMITH CARTON FILLING AND SEALING MACHINE SPEED DEVELOPMENT — —

A 500% INCREASE IN 26 YEARS

1. WHY DO WE NEED QUALITY CONTROL?

In years gone by fairly large variation in the quality characteristics of folding paper boxes were of little significance. Hand filling and sealing, or slow and cumbersome machine methods, allowed fairly wide swings in quality. Today this is no longer true. The graph above shows what has happened in a general way. The super market and new developments in food preservation have greatly expanded the use of the paper

box. Competition and mass production have added the incentive to ever faster packaging machine speeds. These increased speeds have made extremely significant the quality variations which formerly could be neglected. The box maker of today faces the double horns of a dilemma--he must make boxes faster to keep costs down and he must make them better to meet his customers' requirements. The answer is better methods of quality control in the making of the paper board and in the making of the carton.

* * * * * * * * * * * *

WHAT IS STATISTICAL QUALITY CONTROL ?

A MANAGEMENT TOOL —

A SCIENTIFIC APPLICATION OF THE LAWS OF MATHEMATICS TO THE CONTROL OF PRODUCTION QUALITY.

"STATISTICAL QUALITY CONTROL IS LIKE AN X-RAY. IT LETS YOU LOOK INTO A PROBLEM AND SEE THE ANSWERS."

MR. DALE LOBSINGER, Q.C. ENG.,
UNITED AIR LINES.

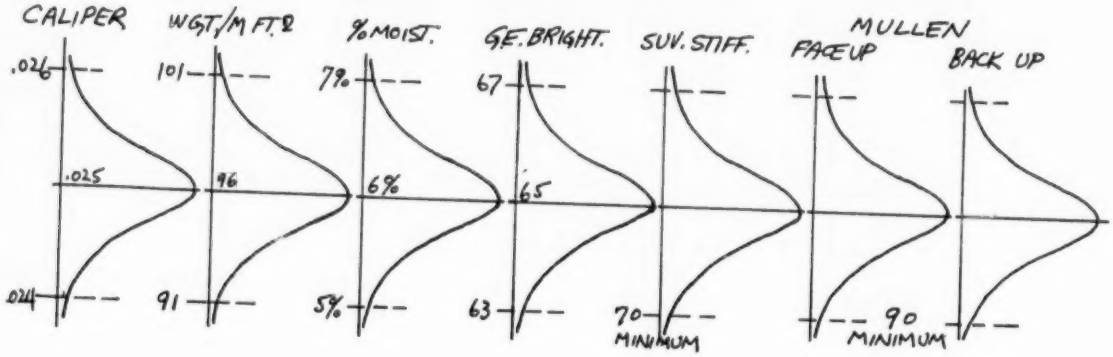
* * * * * * * * * * * *

2. WHAT IS STATISTICAL QUALITY CONTROL ?

To meet the challenge of ever tightening specifications in all industries, a new management tool has been developed. It is named "Statistical Quality Control." Based on a sound mathematical background, this modern method of quality control introduces a new conception of quality in that it frankly recognizes that quality always varies, never is a fixed thing. From this fundamental conception it proceeds to apply known laws and set up limits of expected normal variation. This makes possible the setting of sound scientific specifications based on what the men, materials, and machines involved can produce. By means of an additional technique called control charting, production information is constantly delivered back to production people in a pictorial manner they can understand. Control charts show allowable limits of variation and serve as a quality guide. They make possible the correction of quality trends before they exceed allowable limits. Thus statistical quality control is actually a greatly improved method of communicating quality information to operating people thus making it possible for them to do more and better work.

THE CUSTOMERS SPECIFICATIONS

NO. 1 PATENT WHITE - NEWS BACK MUST HAVE EXCELLENT FINISH - FOR 120 LINE SCREEN HALFTONE PRINTING.

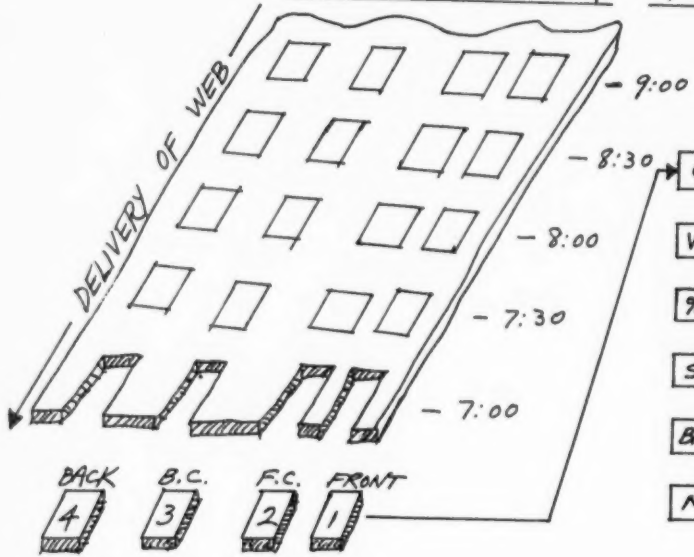


THE CUSTOMER SAYS, "THESE SPECS MUST BE MET!"
CAN YOU MEET THEM?

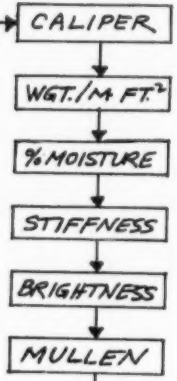
3. THE CUSTOMER'S SPECIFICATIONS

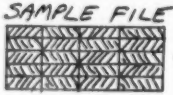
The first step in the production of a good paper box is the making of the box board. Physically speaking, boxboard possesses characteristics of thickness, unit weight, stiffness, tearing strength, bursting strength, folding qualities, and water absortive qualities. In addition, the boxboard is expected to be well formed, uniform in color, free from surface defects, of correct sheet size and having clean cut and square edges. Fortunately, not all of those specifications are often required on any one given order. The customer will usually specify what he wishes and will sometimes state allowable limits of variation. The illustraion at left shows graphically a typical set of these customer's specifications. The problem facing the board maker is "can I meet them?" Here opinions based upon experience are of value, but usually are only approximate, Many customers are now setting a definite maximum percentage of tests which may exceed specifications. Without better methods of setting specifications and determining the process limits, the paper maker faces serious difficulties. Statistical methods will eliminate human opinion and give him the facts he needs.

GETTING THE FACTS



SAMPLING FOR FACTS





TESTING FOR FACTS

INSPECTORS REPORT

CALIPER	7:00	7:30	8:00	8:30
1-FRONT	2250	2525	2500	2500
2	2250	2525	2500	2500
3	2525	2525	2525	2525
4-BACK	2525	2550	2550	2550
TOTAL	10150	10125	10075	10075
AVERAGE	2537	2531	2518	2518
HI TEST	2530	2550	2550	2550
LO TEST	2525	2525	2500	2500
RANGE	25	25	50	50

WGT./M	7:00	7:30	8:00	8:30
1-FRONT	95	96	92	93
2	93	92	92	92
3	89	91	92	90
4-BACK	93	94	92	90
TOTAL	370	373	368	365
AVERAGE	92.5	93.3	92.0	91.3
HI TEST	95	96	92	93
LO TEST	89	91	92	90
RANGE	6	5	0	3

RECORDING THE FACTS

4. GETTING THE FACTS

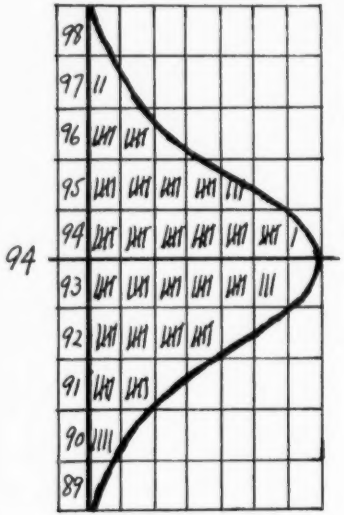
There is always one best way of doing any job. The use of statistical methods of gathering, recording, and analyzing data is the one best way for the paper maker to determine what he can do, and whether or not he can meet his customer's specifications.

One successful method of getting the facts is portrayed graphically at the left. Note that the paper maker produces his product as a continuous web, cuts it (usually) into sheets and sells it by the ton. His customer, the boxmaker, prints the paper by the sheet and sells it by the thousand boxes. Finally, the purchaser of the boxes packages them as individual cartons. The real problem facing the paper maker then is two-fold. First, process control which keeps the general level of quality correct, and secondly, a finer product control which assures the final purchaser that the individual boxes will meet specifications.

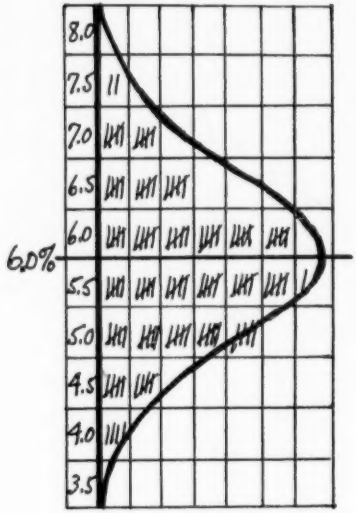
The simple sampling plan shown at the left gives accurate information as to both these quality variables. Variation within sets of four samples gives a picture of individual sample variations and variations between sets of samples gives a picture of general process control. Recording the data as shown at the right makes simple the determination of both the average value of a set of four samples and the difference, or "Range," between the highest and the lowest. With these two bits of information, the average and the range, analysis can be performed which will indicate paper control limits for the averages, and under certain conditions, probable limits of variation of the individual samples. Experience has shown that these analyses give excellent approximations to actual results achieved on the paper machine.

WEIGHT, MOISTURE & CALIPER VARIATION

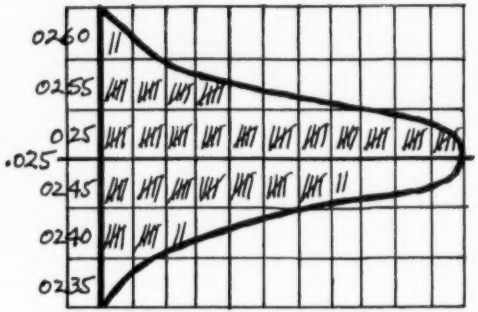
.025 NO. 1 PATENT WHITE-NEWS BK.



WEIGHT
LBS./M FT.²



% MOISTURE
(CHART)



CALIPER

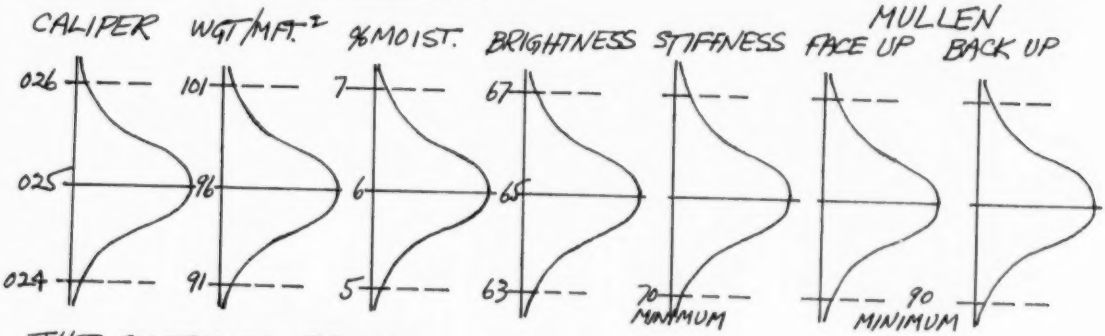
5. WEIGHT, MOISTURE, AND CALIPER VARIATION

If a simple method such as indicated at left (called a frequency distribution) be used to organize the data gathered it will be noticed that the plotted results will generally conform quite closely to a bell shaped curve. This is called a "normal distribution" and indicates that the process was probably operating under "statistical control." That is, quality was varying back and forth in a normal manner. It is possible to analyze further such distributions and determine directly from them the outer limits of normal variation. Ordinarily such methods are too cumbersome, however, and shorter methods are used. These methods require knowledge of the range, or variation between tests, as mentioned in the preceding discussion.

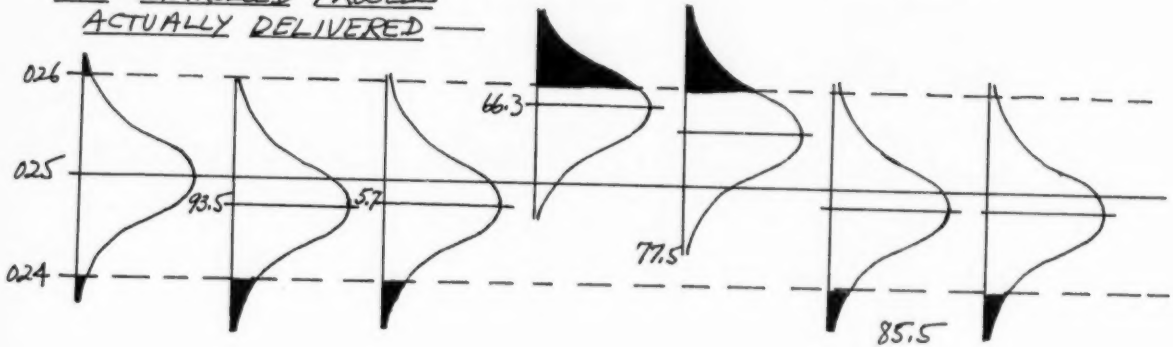
In the above it is important to note the "line up of specifications." It is shown that caliper at .0155 goes with a weight of 63#/M ft. and an average moisture of 6%. These values represent process capabilities and cannot be changed at will. Given a certain quality of paper to be made, a certain materials formulation for it, the papermakers first problem is to determine what this natural "line-up of specifications" is. This is the information that tells him whether or not he can meet customers' specifications, and if not, what he can meet. Armed with this information he can approach the customer with confidence and request a modification of the specifications. Faced with the facts, most customers cooperate wholeheartedly. It has been well said that "Statistical Quality Control substitutes facts for opinions."

PICTURING THE FACTS

THE CUSTOMER SPECIFIED



THE CONTROLLED PROCESS ACTUALLY DELIVERED



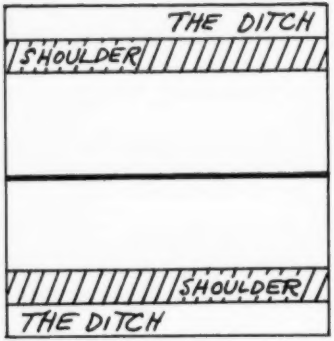
6. PICTURING THE FACTS

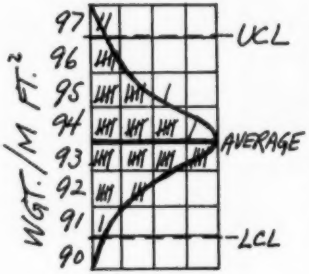
In the figure at left, the upper line of normal distribution curves show the line-up of boxboard specifications requested some years ago by one of our customers. At the time, statistical information was not available, hence the order was accepted with certain reservations. Several re-runs of this order were made over a period of months. Facts were gathered as outlined previously and were analyzed. The bottom row of distribution curves show the actual results achieved. The black areas of the distributions indicate the approximate portions of the various tests that exceeded the customer's allowable limits.

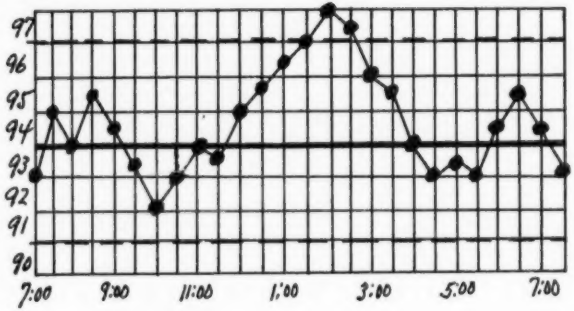
The customer was approached with the facts and readily agreed to modify the specifications to meet the indicated process levels. This was done, and since that time no boxboard complaints have been received from him.

Specifications cannot be determined by edicts from high places. They must be determined by careful analyses of the productive output of the men, materials, and machines available. Statistical methods make such an analysis possible.

THE QUALITY HIGHWAY——







THE QUALITY HIGHWAY

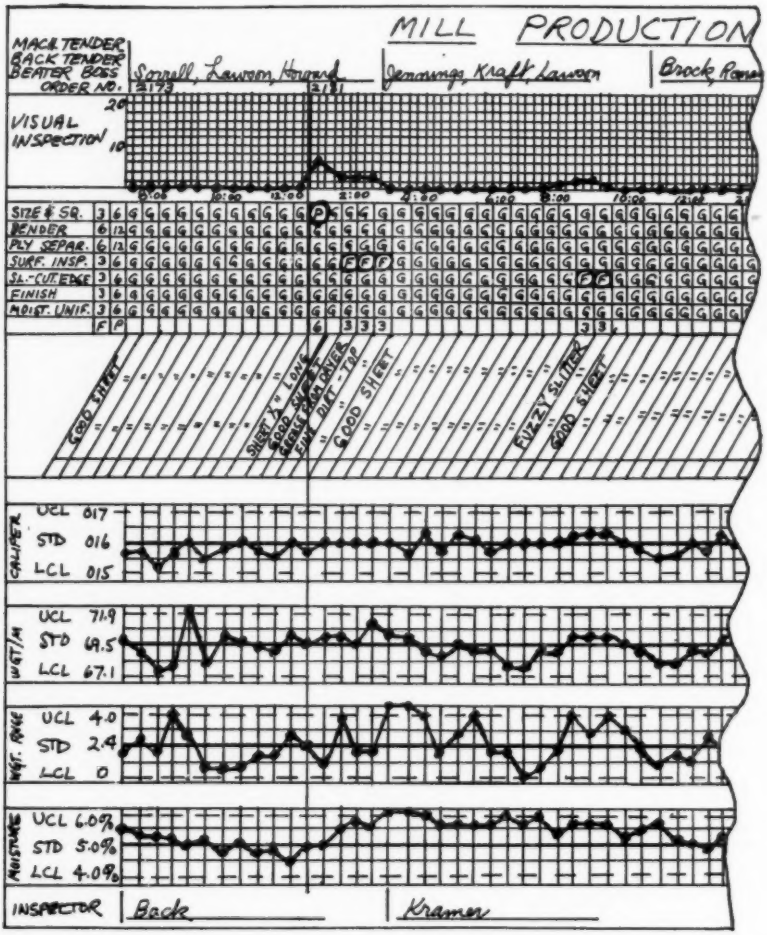
THE CONTROL CHART

7. THE QUALITY HIGHWAY

The control chart is simply a method of delivering operating information to operating people. Dr. Lloyd Knowler, of the University of Iowa, has likened it to a quality highway as illustrated at left. The control limits become the edge of the highway, and specification limits are likened to the edge of the ditch. As long as the quality varies about the center line and remains within the control limits it is said to be "in control." Note in the illustration that the scale of test values appears at the left of the control chart and the time element appears at the bottom. As the tests are run, the average value is plotted as a point on the chart at the proper time interval and quality level.

Note that between 10:00 A.M. and 2:00P.M. a steady trend developed. This should have been caught and reversed before the quality value exceeded the upper control limit. Thus, the two "out of control" points might have been avoided. Graphic information such as this is easily grasped by skilled operating people and serves to guide them, thus reducing out-of-specification production.

BACKTENDERS CONTROL CHART



A. THE PICTURE OF VISUAL QUALITY

B. EVALUATION OF VISUAL QUALITY CHARACTERISTICS

C. DESCRIPTION OF DEFECTS

D. MEASURABLE QUALITY CHARACTERISTICS

8. THE BACKTENDERS CONTROL CHART

The backtenders chart indicates finished product quality. The upper portion, which covers visual inspection, is based upon adaptation of a "c" chart technique. Arbitrary weighting values are assigned to the various quality characteristics as shown at the left. Quality is assumed to fall into three classifications, Good (G), Fair (F), or Poor (P). The inspector is required to record his evaluation of each quality characteristic on each test, score the quality level by adding up the defect weighting values, and plot this value (called the Defective Quality Number) at the proper place on the chart. A system of flagging defective production is also provided.

The bottom portion of the chart shows average caliper, weight per 1000 ft^2., weight range across the machine, and average moisture. These are the most important factors of boxboard quality and are plotted on every order. Other quality characteristics may be plotted on separate charts as required.

Control charts cover a 24 hr. daily cycle and carry names of machine crews and inspectors. This brings about friendly competition between shifts which serves greatly to increase quality and reduce defective production.

TARGET CONTROL LIMITS FOR BOARD MILL CONTROL CHARTS

STD. WGT/M FT.^{2}	90.0	90.5	91.0	91.5	92.0	92.5	93.0	93.5	94
AVERAGE CHART									
UPPER CONTROL LIM.	93.0	93.5	94.0	94.6	95.1	95.6	96.1	96.6	
STANDARD	90.0	90.5	91.0	91.5	92.0	92.5	93.0	93.5	94.0
LOWER CONTROL LIM.	87.0	87.5	88.0	88.4	88.9	89.4	89.9	90.4	90.8
WGT. RANGE CHART									
UPPER CONTROL LIM.	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	
STANDARD	3.09	3.11	3.13	3.14	3.16	3.18	3.19	3.21	3.2
LOWER CONTROL LIM.	0	0	0	0	0	0	0	0	0
MACHINE TENDERS CONTROL CHART									
UPPER CONTROL LIM.	94.0	94.0	95.0	95.0	96.0	96.0	97.0	97.0	9
STANDARD	90.0	90.5	91.0	91.5	92.0	92.5	93.0	93.5	94
LOWER CONTROL LIM.	85.0	86.0	86.0	87.0	87.0	88.0	88.0	89.0	8

THIS STD. WEIGHT FIGURE APPEARS ON EACH MACHINE ORDER

THESE LIMITS FOR AVERAGES OF 4 TESTS ARE USED ON THE BACKTENDERS CHART.

WGT. RANGE FIGURES USED ON BACKTENDERS CHART. UPPER CONTROL LIM. TO NEAREST FULL LB. VALUE. SEE NOTE.

± 5% VALUES FOR MACHINE TENDERS CONTROL CHART. SEE NOTE FOR METHOD OF ROUNDING OFF DECIMALS.

9. TARGET LIMITS IN THE PAPERBOARD MILL

It is not practical to determine control limits for board mill weight/1000 ft^2. by normal methods. Hence target limits are used which are based on Industry standards. It is assumed that individual samples shall be allowed to vary plus or minus 5%. Simple algebraic solutions of standard formulae reduce control limits to terms of $\bar{X}$, the average weight/1000 ft^2., as shown below. Target limits are then determined for progressive values of $\bar{X}$. All machine orders carry specified weight per 1000 ft^2. By simply referring to the proper table, control limits for averages and for individual tests, and also for allowable range values can be quickly determined and put on the control chart before an order is begun. This makes a constant use of control limits possible. The simple algebraic computations (for subgroups of 4 samples) are as follows:

1. Determining allowable $\bar{R}$ in terms of $\bar{X}$.

$$UCL_{ind.} = \bar{X} + \frac{3\bar{R}}{d_2} \quad \text{(ind. = individual sample)}$$

$$\text{or} \quad \frac{3\bar{R}}{d_2} = UCL_{ind.} - \bar{X}$$

$$\text{from which} \quad \bar{R} = d_2\left(\frac{UCL_{ind.} - \bar{X}}{3}\right)$$

$$\text{By definition} \quad UCL_{ind.} = (\bar{X} + .05\bar{X}) = 1.05\bar{X}$$

$$\text{hence} \quad \bar{R} = d_2\frac{(1.05\bar{X} - \bar{X})}{3}$$

$$\text{simplified,} \quad \bar{R} = \frac{2.059 \times .05\bar{X}}{3}$$

(where $d_2 = 2.059$ for sample size four)

Finally, by simplification:- $\bar{R} = .0343\bar{X}$ & $UCL_R = .0782\bar{X}$

2. Determine control limits for average chart based on $\pm 4\sigma$ variation (an arbitrary limit based upon past experience which covers actual conditions pretty well).

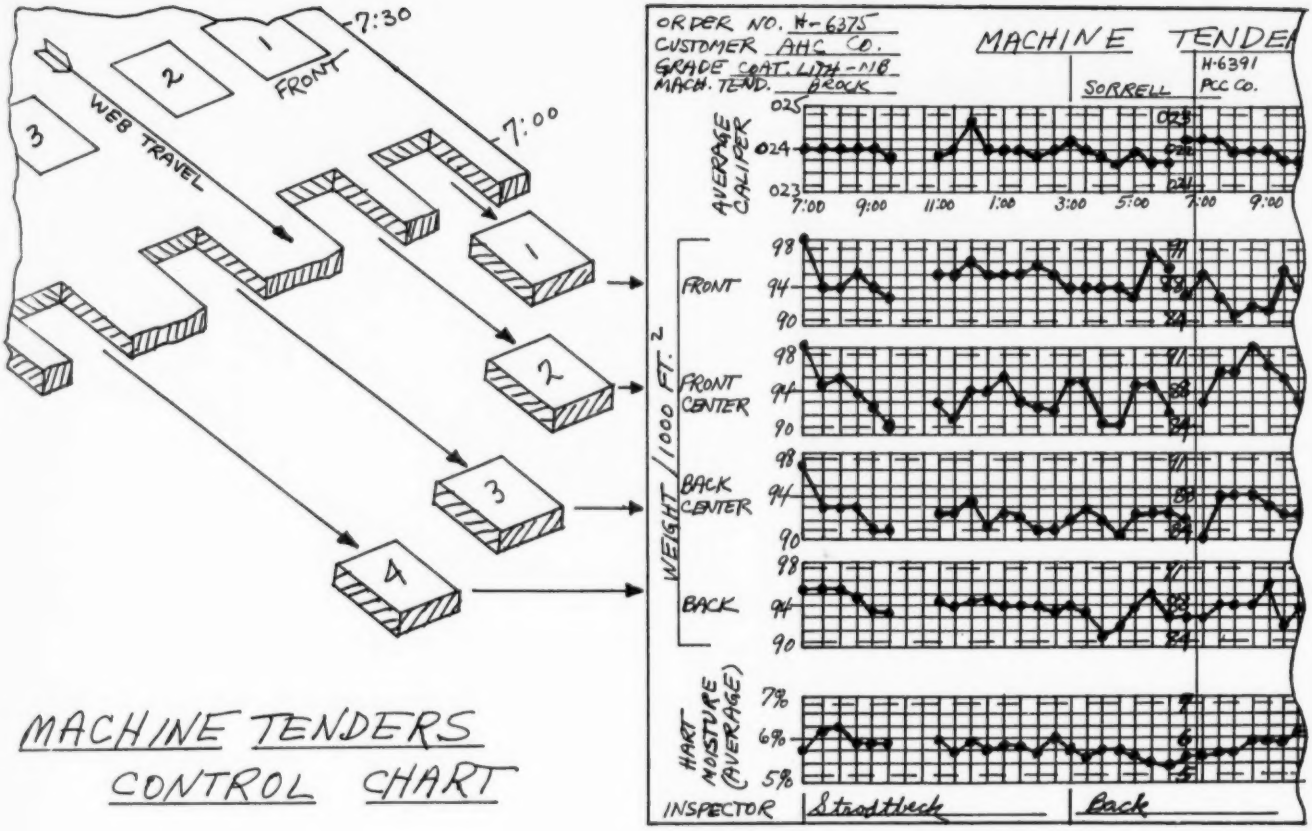
$UCL = \bar{X} \pm A_2\bar{R}$ for 3σ limits & $\bar{X} \pm \frac{4}{3}A_2\bar{R}$ for 4σ limits

but $\bar{R} = .0343\bar{X}$ determined, and since $A_2 = .73$ for sample size four, therefore, by substitution —

$UCL = \bar{X} \pm (\frac{4}{3} \cdot .73)(.0343\bar{X})$, from which by simplification—

$$UCL = \bar{X} \pm .0334\bar{X} \quad \text{and} \quad LCL = \bar{X} - .0334\bar{X}$$

This reduces all control limits to terms of $\bar{X}$, and simple calculations make the table above possible.



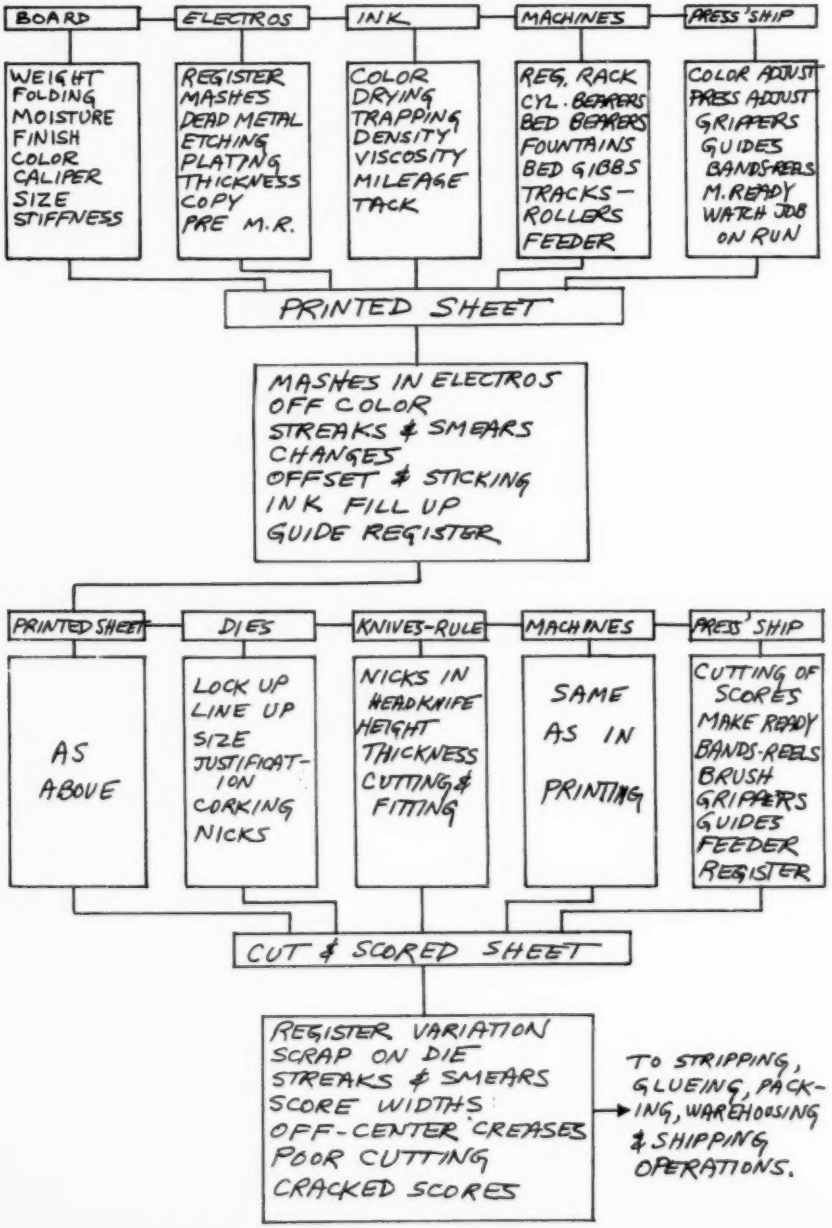
MACHINE TENDERS CONTROL CHART

10. THE MACHINE TENDERS CONTROL CHART

The machine tender requires slightly different information. It is his responsibility to maintain control over the process, therefore he needs detailed information rather than just average values. Experience has shown that if uniform weight and density can be maintained across the machine, other tests will generally fall pretty well in line. Hence, the machine tender wishes to know how the weight is running at different places across the width of the machine. Since the inspector has already obtained this information, it is a simple matter to set up a series of graphs that visually portray it. This has been done, and a quick glance will tell the machine tender where he stands with respect to standards at any one of the four positions across the machine. He is also interested in average caliper, and moisture, and these are also shown. Use of this chart on one machine reduced variation across the web by nearly 50% in six months. There is no substitute for operating informationdelivered promptly to the operating personnel that can use it.

The foregoing gives only an introduction to statistical quality control methods as they can be used in paper board mills. Many other applications are possible and the same general approach can be used. There is no comparable method for determining facts, and once process limits are determined and proper control limits set, improved quality follows as a matter of course. It must be remembered, however, that quality control methods cannot <u>put</u> quality into the product. That must be built into the product by operating people. Statistical Quality Control is only a <u>tool</u> for management to use, but it has been termed as the most valuable tool produced in the last half century.

SOME SOURCES OF DEFECTS IN THE PRODUCTION OF FOLDING CARTONS



11. SOURCES OF DEFECTS IN THE PRODUCTION OF A FOLDING PAPER CARTON

The chart at the left shows a few of the sources of trouble and defects in the production of folding cartons. True Quality Control must begin with the basic raw materials. We have covered the control methods for paper boxboard, and similar methods can be used to control the quality of electros, inks, and machine operation. The quality of the human element, workmanship, must be developed by line management.

Let us assume we have the basic factors of good men, materials, and machines. The combination of these factors in the printing operation gives rise to other defects, while successive steps of production create still others. True quality control seeks to control quality at these basic steps of production. Hence control stations are set up to check each one. In the carton plant this means three major stations - in printing, cutting, and creasing, and in gluing operations.

STATISTICAL BACKGROUND

1. The Poisson Distribution applies when an event has many opportunities to occur but is extremely unlikely to happen at any given single opportunity. This describes quite accurately the conditions under which most carton fabrication defects occur, and is the mathematical model used.

2. The Defective Quality Number - an index to the quality level of a product produced under conditions subjecting it to many different kinds of defects, any of which may vary in severity from minor to critical.

$$DQN = \sum(dnr) \times \frac{AX}{N}$$

d = name of defect
n = number of defects d
r = severity number
A = area of sheet in ft^2.
X = number of times thru press
N = number of units on sheet

3. The Control Limits - In the interest of inspection efficiency, some freedom is taken with simon-pure statistical theory in determining control limits.

$$\begin{matrix} U \\ L \end{matrix} CL = \overline{DQN} \pm 3\sqrt{\overline{DQN}}$$

From this form a simple tabulation of UCL and LCL values can be derived for given values of DQN.

4. Reducing Visual Defects to Numerical Values

Defects are -	Severity No. is -
CRITICAL - when the product is rejectable	6
MAJOR - when the product quality is of questionable acceptance by the customer	3
MINOR - when quality is slightly off standard but still acceptable to the customer	1

* * * * * * * * * * * *

"IF QUALITY CAN'T BE MEASURED - IT IS'NT QUALITY."

Dr. Lloyd Knowler
University of Iowa

12. STATISTICAL BACKGROUND

Some liberties are taken with simon-pure statistical procedures. This is done deliberately for the purpose of keeping the procedure simple and making it possible for inspectors to spend a bare minimum amount of time on clerical work.

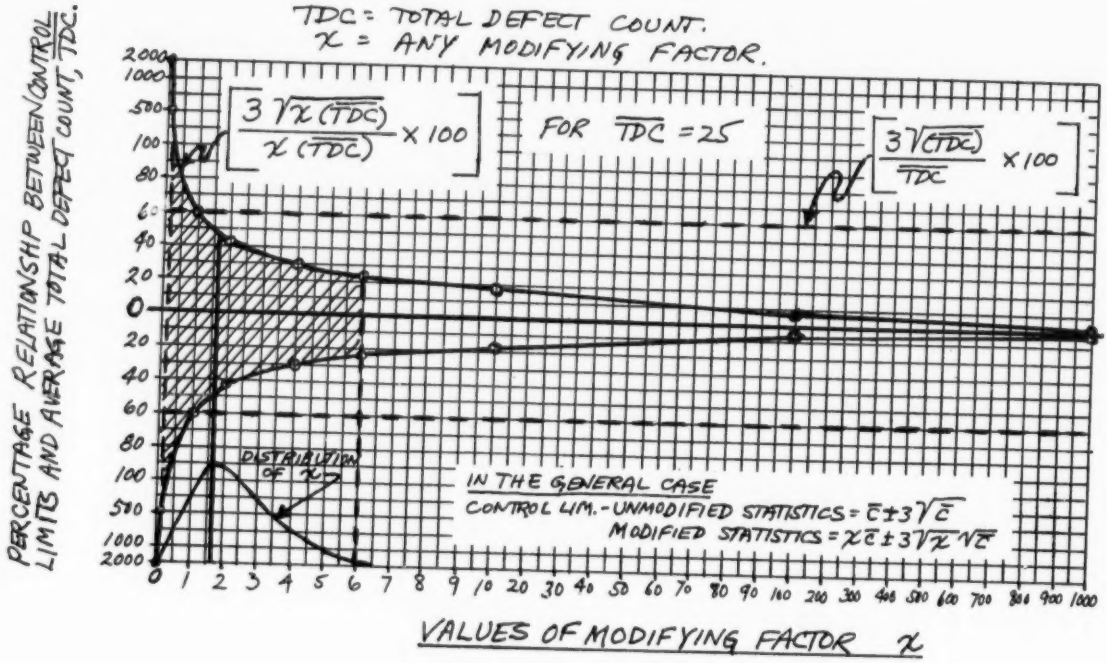
The Poisson distribution most closely approximates the conditions encountered in paper carton fabrication and is the mathematical model used. This makes possible a modified "c" chart technique, which is fortunate, for it is simply handled.

Defects are considered as falling in three categories - minor, major, and critical. Weighting values of 1,3, and 6 respectively are assigned to these degrees of severity and are called severity numbers. Other values could as well be used. These have proven satisfactory. The DQN, or Defective Quality Number, is simply a summation of the severity numbers for the defects found at a given inspection, multiplied by a modifying factor which serves to key all different jobs into a common level on the control chart. The modifying factor $\frac{AX}{N}$ is computed once for a given order and then becomes a constant in the formula for DQN. Computations for the DQN for each inspection is a very simple matter and takes but little of the inspectors time.

Control limits are simply determined and, in fact, are tabulated in a booklet for progressive values of DQN. The inspector needs but to determine the average DQN - ($\overline{DQN}$) after running some 15 or more inspections and then read the control limits out of the booklet.

Broadly speaking, the statistical background is sound, In the interest of "on-the-job" efficiency, some short cuts have been effected, and have had but little effect upon the program. Greater statistical accuracy could have been had at a cost of reduced inspection -- a poor bargain.

FIG. 12-B - RELATIONSHIP BETWEEN C CHART CONTROL LIMITS DETERMINED FROM MODIFIED AND UNMODIFIED STATISTICS.


TDC = TOTAL DEFECT COUNT.
χ = ANY MODIFYING FACTOR.
PERCENTAGE RELATIONSHIP BETWEEN CONTROL LIMITS AND AVERAGE TOTAL DEFECT COUNT, TDC.
$\left[\frac{3\sqrt{\chi(\overline{TDC})}}{\chi(\overline{TDC})} \times 100\right]$
FOR $\overline{TDC} = 25$
$\left[\frac{3\sqrt{(TDC)}}{TDC} \times 100\right]$
DISTRIBUTION OF χ
IN THE GENERAL CASE
CONTROL LIM. - UNMODIFIED STATISTICS = $\bar{c} \pm 3\sqrt{\bar{c}}$
MODIFIED STATISTICS = $\chi\bar{c} \pm 3\sqrt{\chi}\sqrt{\bar{c}}$
2000 1000 500 100 80 60 40 20 0 20 40 60 80 100 500 1000 2000
0 1 2 3 4 5 6 7 8 9 10 20 30 40 50 60 70 80 90 100 200 300 400 500 600 700 800 900 1000
VALUES OF MODIFYING FACTOR χ

THE USE OF CONTROL LIMITS BASED ON MODIFIED STATISTICS

The use of control charts on any printing operation presents two unique problems. In the first place, many kinds of defects can occur, any one of which can vary in seriousness, from being minor and of no consequence to being critical and forcing complete rejection of the product. When a sheet is inspected and the TDC (total defect count), or the $\Sigma (d_\chi n_\chi r_\chi)$ as it has heretofore been defined, is determined, the TDC can be regarded as an unmodified statistic and, handled in a standard c chart manner, the control limits will be valid. However, since the TDC is composed of a summation of three separate classes of defects a point outside the control limits may not necessarily indicate a bad quality condition. Let us illustrate: Assume a given sheet having 24 boxes on the sheet. Assume the TDC over a period of time to average 9.0, in which case the upper control limit would be $9 + 3\sqrt{9}$ or 18.0. Upon inspection it is found that all boxes on the sheet are slightly out of register due to the sheet missing a guide just by a few thousandths of an inch. The defect is minor (No. 1 severity), and the quality is still good. The TDC is 24 x 1 = 24. The point would lie outside the upper control limit and would rightfully indicate an assignable cause of the defective work. From a practical standpoint it would not force a shutdown of the press, but would warn the pressman to carefully watch the job to see if the condition persisted. Now suppose on the next inspection the general register is OK, but one printing plate is found to have slipped badly (No. 6 severity) and the box is therefore rejectable. Now the TDC is 1 x 6 = 6 and the point plotted is well with in the control limits, even being less than the average. However, action must be taken immediately to minimize spoilage, How this information is brought to the pressman's attention will be explained later.

At first thought, it appears that control limits in such cases are of minor importance or even misleading. However, note that on the overall picture, the control limits give a picture of quality and do serve their normal function. Actually, such instances as the illustration very rarely occur, serious defects almost always being accompanied by minor ones. In this first unique case then, control limits may or may not indicate the presence of seriously defective work, but do serve a very useful function in giving a picture of average running quality.

The second unique problem is the wide variation of economic value. Every printing job presents different economic factors. The quality of paper varies, as does sheet size and type of printing. The number of colors may vary from 1 to 6 or more with an ever increasing ink, labor, and press time cost. Therefore the circumstances under which a defect occurs have a great deal to do with its importance.

Again let us illustrate: Assume a sheet size 36 x 48 with 12 boxes on the sheet. Now, as before, let us assume that one printing plate slips out of register and the box must be thrown away. 1/12 or a little more than 8% of each sheet so run is spoiled. Now consider the same size sheet having the same defect on it, but having only two boxes on the sheet. Here 50% of each sheet must be rejected with a much greater loss of materials, etc. One can easily appreciate that the same defect is much more important in the second case than in the first. The circumstances have determined its importance.

To care for these varying circumstances, the total defect count, or TDC, is modified by an economic factor which is determined by the

conditions of the job. This factor $\frac{4x}{4}$ has been discussed previously, and is determined by multiplying the area of the sheet in square feet by the number of times through the press and dividing the answer by the number of units on the sheet.

Control limits determined from modified statistics differ greatly from those determined from unmodified statistics. Note the comparison of the basic formulae for the general case,

1. $CL = c \pm 3\sqrt{c}$ (unmodified statistics)
2. $CL = x(\bar{c}) \pm 3\sqrt{x}\sqrt{\bar{c}}$ (modified statistics)

In 2 the average varies as a direct function of x, the modifying factor, while the control limits vary as the square root of the modifying factor. A picture of the effect of this variation is shown in Figure 12B. Note its calculated effect.

As the economic factor increases above 1, the control limits narrow, thus emphasizing the importance of defects and calling attention to them more quickly. This is exactly the result desired. As the economic factor decreases below 1, the relative importance of given defects decreases and this is recognized by widened limits. Handled in this manner, control limits become target limits, the target being determined by the economic factors of the job. It is true that limits so set insure a certain percentage of out of control points. However, since the control limits are not the basis upon which action is taken, this has presented no serious problems. They do serve as a guide, and being narrowed, do emphasize the importance of defects. The value of the practical effect achieved has more than outweighed the importance of the theoretical error involved. In all cases, serious or critical defects are brought to the attention of operating crews as will be discussed later (Figure 16, The Signal System).

These two unique problems--the use of a single control chart to cover three different degrees of severity of many diffenert kinds of defects, and the constant variation of the economic factors which determines the importance of a defect -- have made necessary the use of two somewhat new techniques. First, the use of control limits based on modified statistics and second, an auxiliary method of warning operating crews of the presence of serious or critical defects. Both techniques have proven very effective in controlling printing quality.

FIG. 13 PRINTING DEPT. CONTROL CHART—

PRINTING DEPARTMENT
DEFECTIVE QUALITY CONTROL CHART

PRESSMAN Cogan

ORDER NO. M-5393
CUSTOMER ALISTAIR CO.

REMARKS (INITIAL ALL NOTATIONS)

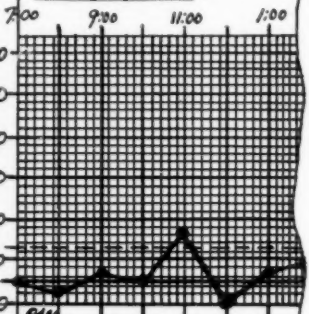
11:00 a.m. Off register Boxes 3 and 4 OK for #1 severity. A.H.L.

#6 mach on box No. 3 at 11:00. Ran 200 defects. Pressman flagged load. B.L.

☑ 1ST RUN
☐ 2ND RUN
☐ 3RD RUN
☐ 4TH RUN

$\frac{AX}{N} = 1.6$

$N = 8$



COUNTER READING	6430							
DEFECTS - THIS PRESS								
OFF REGISTER				1X1	2X1	2X1		
OFF COLOR								
MASHED PLATE						(1X6)		1X1
INK FILL-UP		1X1	(1X3)					
DIRTY FORMS					1X1	2X1		3X1
SMEARS & STREAKS								
SIDE GUIDE MARK	OK	OK	OK	OK	OK	OK	OK	OK
DEAD METAL	3X1	1X1						
DEMERITS TOTAL	3	2	4	3	10	0	4	
NOTES ON DEFECTS	DM.-2-4B	FILL UP-2 DEAD METAL-7	FILL UP-3 O.REG.-3	O.REG.-3-4 DIRTY FORM-7	MASH-3 O.R.-3-4 DIRTY FORM-7		MASH-6 DIRTY FORM-7	
DEFECTS OKAYED BY FOREMAN								
OFF REG.-OK BY FOREMAN						2X1	2X1	

INSPECTOR Lukens

Notes are valuable, especially in cases of trouble. Inspectors and Foremen are encouraged to use the "Remarks" space.

The graph line gives a good picture of quality. Since each point represents a "Defective Quality Number," the aim is to keep as near zero as possible.

A few common defects (the first 6 listed) account for most of the defects. Others are written in as they occur.

The "notes on defects" tell the pressman where the trouble is. Being able to locate it quickly saves down time.

Pressmen do not like to be charged with defects they cannot correct. If they can convince the foreman this is so, he will O.K. the defect. The inspector will then record the defect on the bottom part of the chart and it is no longer charged against him.

13. THE PRINTING DEPT. CONTROL CHART

Quality Control is obligated to make the production of quality products as simple as possible. The D.Q.N. control charts have been designed to do exactly that. The charts are located at each press, and following each inspection, the inspector goes immediately back to the press and posts the information on the chart. The graph line gives a visual picture of quality. Immediately below the graph the defects found are listed both as to number and severity thereof. The 2x3 appearing in the column for "Dirty Forms" means two boxes defective with major (#3) severity. According to custom in the trade the boxes on the sheet are numbered according to a fixed plan. At the bottom of the chart in the section "Notes on Defects" the inspector records the defects found and the box number defective. This makes it possible for the pressman to shut down the press and make needed repairs with a minimum of lost time. As in the board mill, the charts are 24 hour charts and tend to induce friendly quality competition between operating crews.

In both theory and fact the charts are designed to help press-men, not police them. As a result they are quickly recognized as a help and pressmen give good support to the program. After posting control charts in his pressroom for six months one Superintendent declared his printing to be at the highest level in twenty-five years.

CUTTING AND CRE

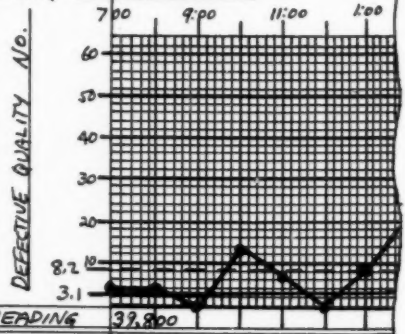
DEFECTIVE QUALITY

PRESSMAN Jones

REMARKS (INITIAL ALL NOTATIONS)	
Off register - Box 3,7 OK for #1 severity JRC	
Poor Color. 9:00 - 10:00 Light on Blue. Leads flagged by pressman. Anders	

ORDER NO. M-5393

CUSTOMER ALISTAIR CO.



$\frac{AX}{N} = 3.2$

$N = 8$

COUNTER READING	39,800						
DEFECTS - THIS PRESS							
OFF REGISTER							
DEFECTIVE SCORES							
STUCK PANELS		1X1		3X1	2X1		1X3
POOR CUTTING	1X1			1X1			
LARGE NICKS							
SCRAP ON DIE							
GLOSS INK	Good	Good	Good	Good	Good	Good	Good
COLOR	Good	Good	Poor	Poor	Good	Good	Good
DEMERITS TOTAL	1	1	OK	4	2	OK	3
NOTES ON DEFECTS	P.C. - #4	S.P. - #1		P.C. - #8 S.P. - #1,2,7	S.P. - #3,8		S.P. - #3
	DEFECTS OKAYED BY FOREMAN						
OFF REG. - #3,7	2X1	2X1	2X1	2X1	2X1	2X1	2X

INSPECTOR Anders

Defects listed cover almost all defects encountered. On the actual chart, three blank spaces below "color" are provided so that other defects can be written in as they occur.

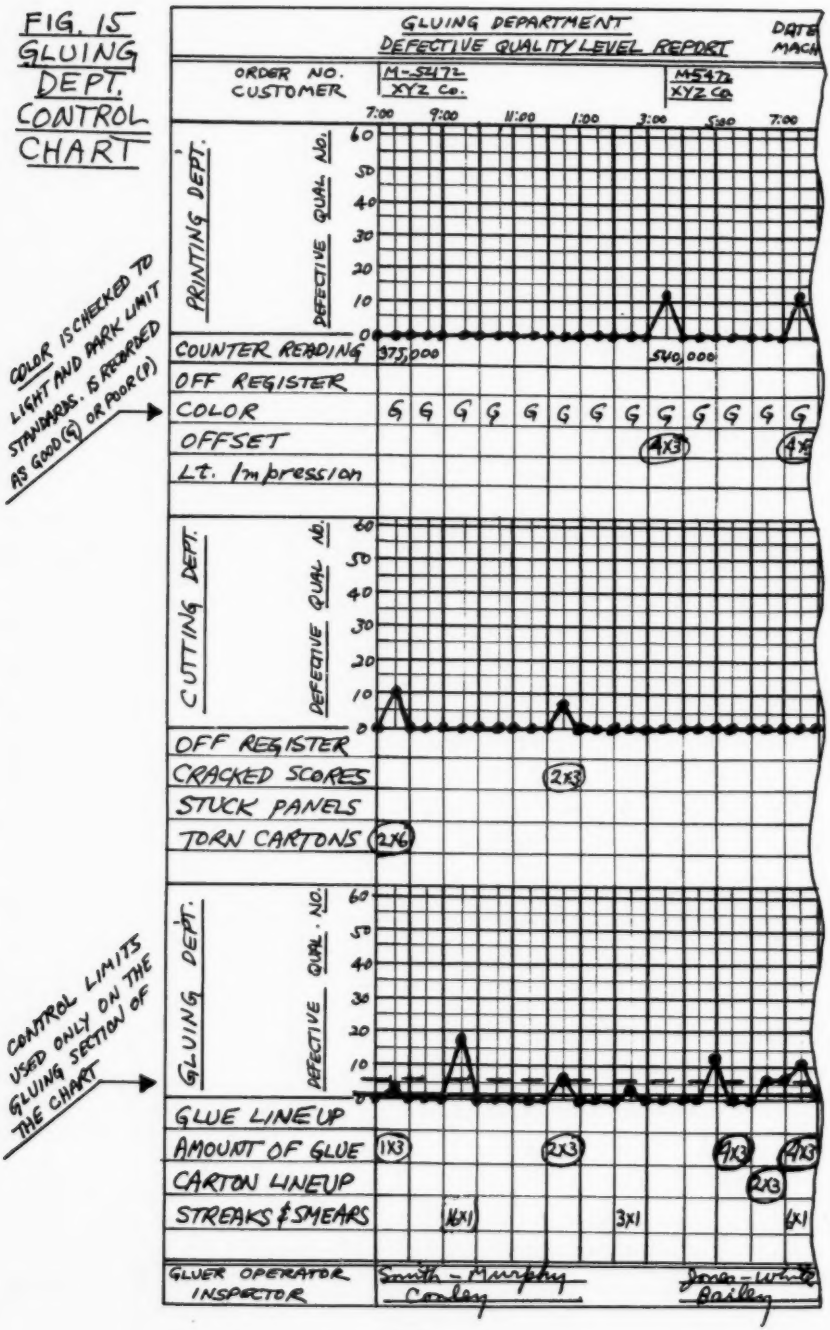
"Color" is checked to light and dark limits provided by the Sales Dept. Within limits color is called "good" and if it is lighter or darker than limits it is called "poor." No effort is made to check color in the Printing Dept., for girls do not have the necessary experience to check wet inks against dry samples. This must be done by the foremen and pressmen.

14. CUTTING AND CREASING DEP'T. CONTROL CHART

In theory and fact this chart is exactly the same as that used in the Printing department, with the exception that the defects themselves are, of course, different. In practice, inspectors are trained to inspect in all departments and the use of identical theory tends to unify the over all procedure.

It will be noted that gloss ink and color are checked in the cutting department. Results are recorded simply as good, fair, or poor, and of course are not charged up to the cutting operation since they occurred in the printing department. Offset, or the transfer of ink from the face of the sheet to the back of the sheet above it is also checked on the same basis.

FIG. 15
GLUING DEPT. CONTROL CHART



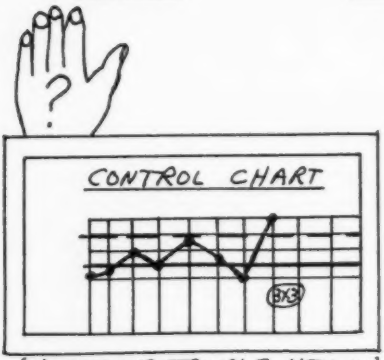
15. GLUING DEPARTMENT CONTROL CHART

The problem in the gluing department is somewhat different than in the others. The gluing operation affords the last opportunity for inspections prior to shipment to the customer. Here the inspection is designed to serve as a check on outgoing printing and cutting quality as well as the gluing operation itself. The top portion of the chart provides a place to record printing defects which have slipped through. The middle portion is for cutting and creasing defects. In both of these cases, only serious or critical defects are checked and recorded. The minor defects are of no consequence and are not even noted. The bottom portion of the chart is used as a control chart for the gluing operation itself, and here minor defects are noted and recorded, for they may be of control value.

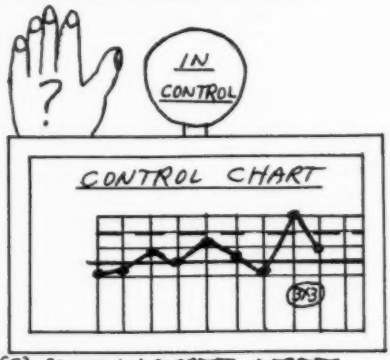
This chart is kept right at the machine, as are the charts in the other departments. This chart serves, as will be noted, a dual purpose. First, a control chart on the gluing operation, and secondly, to provide a picture of the outgoing quality level on the finished cartons. The last factor is of value in determining whether or not sorting is necessary, and if so, to what extent.

Control limits are used only on the bottom portion of the chart. In practice, the inspector will handle about three machines, and will go directly from one machine to the other. Chosen by the inspector at random, 16 cartons are taken one or two at a time just before they are placed in the shipping containers. It takes about seven to ten minutes to inspect and record a test. Serious defects are plainly marked on the cartons which are left beside the control chart by the inspector when she records her findings. These are of value to the machine operators as a guide to making the necessary machine adjustments to correct the defects.

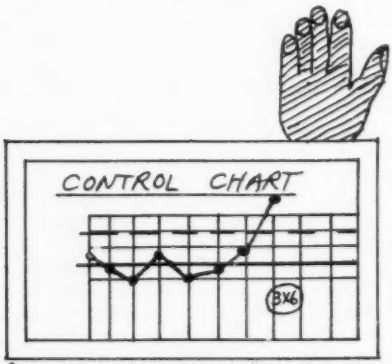
FIG. 16 - - THE SIGNAL SYSTEM



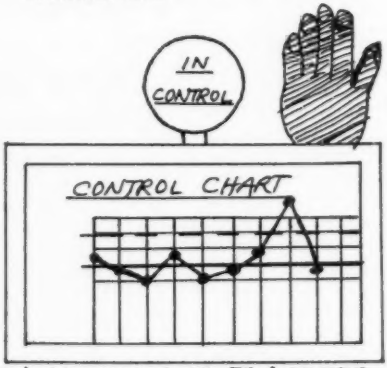
(A) SERIOUS TROUBLE - YELLOW HAND UP - "IN CONTROL" DOWN.



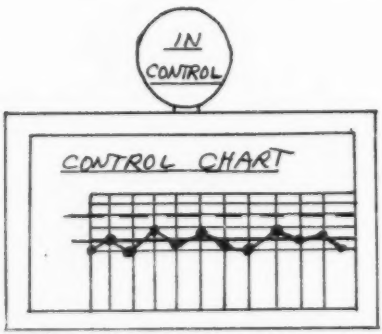
(B) PRESSMAN CORRECTS DEFECTS, PUTS "IN CONTROL" UP - RESUMES PRODUCTION.



(C) CRITICAL TROUBLE - RED HAND UP - "IN CONTROL" DOWN.



(D) PRESSMAN CORRECTS DEFECTS, PUTS "IN CONTROL" UP - RESUMES PRODUCTION.



(E) RUNNING OKAY

16. THE SIGNAL SYSTEM

It is not advisable to have inspectors telling skilled pressmen how to run their jobs. However, if something of a serious nature is found by the inspector, this information should be gotten to the pressmen as quickly as possible to minimize defective work run. This is simply accomplished by the signal system shown at the left. Located behind the chart in such a manner that they can be rotated up or down as needed, are these signals. On the left is a hand painted bright yellow with a black question mark on the palm. In the center is a circular green signal upon which are the words "In Control." On the right is a bright red hand having "Out of Control" printed on the palm. Everyone is acquainted with the color scheme - yellow is caution, green means go, and red means danger.

As long as the job is running normally and no serious or critical defects are found, the inspector will keep the green "in control" signal up as shown at E. Should a serious (but not critical) defect or defects be found the inspector will rotate the green "in control" signal down and will rotate the yellow caution hand up, as at A. This immediately notifies the pressman that something is wrong that needs attention. At the same time the inspector records the information concerning the defect on the control chart. Without having said a word, the inspector now has told the pressman, a) something serious is wrong, b) what the trouble is, c) how serious the trouble is, d) how many boxes are defective, e) which boxes on the sheet are involved.

After correcting the defect, the pressman is not permitted to put down the signal hand but he is permitted to put up the green "in control" signal as shown in B and proceed with production. The signal system now, in effect, says, "We had trouble, we found it, we fixed it, and we are running Okay again."

Exactly the same method is used for critical defects except the red hand is used as at C and D instead of the yellow one. The system is simple and effective, and unfailingly accomplishes its purpose - to get presses shut down when defective work is being produced.

A DAYS RESULTS

As a brief picture of how Quality Control is functioning in one carton plant, note the following data from a typical days run:

Inspections in Printing Dept.	157	3 shifts
Inspections in Cutting Dept.	90	3 shifts
Inspections in the Gluing Dept.	128	2 shifts
In all	375	Inspections

This is an average of 37.5 inspections per 8 hours for each inspector, or one every 12.8 minutes. During this time, the following number of defects were found and immediately reported:

	MINOR	SERIOUS	CRITICAL
Printing	576	7	0
Cutting	97	6	1
Gluing	24	39	16
	697	52	17

The gluing defects listed also include serious and critical printing and cutting defects carried over from previous operations and found by the gluing inspector.

These findings represent vital operating information which was immediately made available to operating crews by the control charts, thus minimizing defective work run.

The above we accomplished at a very low cost per inspection, and undoubtedly saved many times the total cost in reduced defective cartons.

These simple facts point out very well that Quality Control is "an extra pair of eyes" for foremen and operating people.

THE QUALITY PROBLEM

WHAT IS QUALITY?

ACCEPTABILITY TO THE CUSTOMER

WHERE DOES QUALITY COME FROM?

MEN MATERIALS MACHINES

WHAT IS THE PROBLEM?

TO PROVIDE THE QUALITY THAT IS NECESSARY TO SOLVE OUR CUSTOMERS PROBLEMS.

* * * * * * * * * * * *

WHAT IS QUALITY CONTROL'S RELATION TO STANDARDS?

1. TOP MANAGEMENT - SETS STANDARDS
2. LINE MANAGEMENT - ADMINISTERS STANDARDS
3. QUALITY CONTROL - MEASURES CONFORMANCE TO STANDARDS

CONCLUSION

A quality control system has one major obligation--to make the production of good quality as easy as possible for the machine crews. When carefully analyzed, it is noted that the whole operation is simply a streamlined method of getting production information to the man on the firing line while it is news rather than history. It actually turns out to be an extremely efficient communications system between inspection and production. This is as it should be.

In designing such a system, the machine operators viewpoint must be kept constantly in mind. What does he want to know? How does he want it told to him? Are his rights, his pride and his authority preserved? Success or failure will hinge largely upon how these points are handled.

Men, materials, and machines make quality. If we wish to improve quality, one or more of these factors must change. Long experience has shown that people change slowly. Therefore patience is always required, and the long viewpoint is much more in order than one expecting immediate and startling results. The constant pressure of facts as shown by control charts will have its effect, but it is a gradual effect somewhat like the turning of a great ship at sea -- slow but sure.

The ultimate success of such a program also depends in a great degree upon the support given by management. Workmen will place no more importance upon a program than does their supervision. Therefore, both top and line management must stand squarely behind the program.

If these things are done, and if problems are settled one by one as they occur, a Quality Control program may be expected to produce good results through the hands of others, for of itself it can do nothing. It may be trite, but it is true that "Quality cannot be inspected into a product -- it must be built into it."

MANAGEMENT'S USES OF QUALITY CONTROL

Charles R. Scott, Jr.
Otis Elevator Company

Less than ten years ago, as we all remember well, the first atomic bomb exploded into one of the most important events in the history of what we call "civilized man." We witnessed at that time the actual occurrence of something which seemed to most of us to be straight out of the realm of science fiction, an incredible event which required that we rearrange our thinking in order to comprehend what had happened and what it would mean. We are now only too thoroughly aware of the changes in our thinking and feeling which have resulted from that first explosion. Our thinking tells us that the world has overnight shrunk in size; our feelings tell us that it is no longer as safe as it was.

Many other things have contributed to this change in our thinking: the war-time use of radar and penicillin and the tactical use of extremely high-altitude rocket-powered missiles represented rapid strides in the application of science to practical problems. These and many others were precursors of the post-war era during which time the rapid advance in scientific achievement which characterized the war years has extended itself into the greatest period of peace-time achievement in the history of man. Sitting in the middle of this period as we are, we tend to lose sight of the overall trend, but if we stop to compare the period immediately preceding the war with the things going on around us today it becomes clear that we are entering--are already launched upon--an entirely new and different kind of age, the age of Scientific Advancement. Commonplace events of today were, ten years ago, confined to the realm of pure fantasy. It is no novelty today to hear people on the street discussing problems of space travel or future applications of atomic energy, or to read newspaper articles about new experiments in extracting minerals from the sea, raising crops from barren and wasted ground, curing age-old diseases with so-called "wonder-drugs." Truly Science has moved out of the laboratory and into our everyday lives. And the most important aspect of this new emphasis on scientific application is that each new device, each successful new experiment becomes part of a picture which, although still vague and unformed in many places looks very much like a completely new and vast frontier of still unexplored scientific wealth. This picture suggests that we consider the future potentialities of these new servants of man: atomic power, electronics, modern chemistry and others.

I have been speaking in general terms of the foundations of a new age; of more immediate concern to all of us is how this change will affect the industrial processes, particularly where will quality-control fit in. The application of scientific methods to industry is much older than the harnessing of atomic power; it has been in evidence throughout the development of industry. More recently we recall the names of men like Taylor and Gilbreth, Gantt and Shewhart. But today we can discern a new emphasis on scientific methods of manufacture, a sudden spurt in the application of new and exact ways of making things better, faster, and at less cost. The mass production of interchangeable parts associated with American industrial methods is leading the way in adapting an ever-growing percentage of the manufacturing process to automatic methods. Human judgment is becoming a smaller and smaller determinant of the quantity and quality of finished goods.

We can visualize a time in the future when entire factories will be automatic, with skeleton maintenance crews performing the only manual labor. In fact, at least one such plant is already in existence in this country, 97% automatic, with even inspection and control functions performed by computers and machines. This is the type of thing that today's industrial trends seem to forecast. And the development of Quality-control into an exact industrial method ranks high in importance among these trends.

The inspection functions in industry, formerly so much dependent on human ability and discretion, are now being rapidly mechanized into an exact science. The profilometer and surface analyzer, spectrophotometer and various Magnaflux devices are replacing the micrometer, go-gage, and other hand tools; completely new test fields such as industrial radiography and radium and gamma-ray test are being evolved, and the application of statistical methods throughout is being continually improved. As a result of these things quality of goods is being constantly bettered while at the same time inspection costs are dropping off.

We can state with some assurance at this time that Quality-control has already passed through the experimental stage and has now become an accepted part of industrial technology. True, there have been many failures in the past but these are not important; what is important is the number of successful applications which have been made because these prove beyond doubt that, properly managed, Quality-control can be a most effective device for achieving better products at lower costs. We can look forward with confidence to an increasing acceptance, an expanding application of tested methods, and a steady improvement of existing techniques.

Let us look for a moment at some of the factors which are working toward wider acceptance of Quality-control procedures. One such factor and an exceedingly strong one is the attitude of government toward Quality-control in the production of defense materials for government contracts. Quick to appreciate the advantages of Quality-control methods and results, the armed forces have done an excellent job of spreading the doctrines in industrial circles. In particular the Air Force acceptance of QC tools is a milestone in the development of Quality-control as an industrial force, for perhaps in no other field is personal safety so closely tied to the proper functioning of so many machines parts. To those of us connected with production for defense needs references to government quality-control specifications in terms of MIL-Q-5923, MIL-STD-105A, etc., have become almost "household words." I recently interviewed an Air Force Quality-control representative who made a remark which is interesting in this connection. He said, "I don't know where you could get a better ready-made quality-control specification than MIL-STD-105A for thirty-five cents."

We may also take briefly into consideration the recent agitation for the removal of barriers to world trade which seems to be gaining ground rapidly--the "Trade--Not Aid" movement. It seems highly probable that at least some relaxation of tariff walls is bound to result and that American imports and exports will both rise substantially above present levels. This will mean an unprecedented competition between American and foreign-made goods with new premiums being placed on higher quality and lower production costs. We have every reason to believe that Quality-control will play an increasingly important role in achieving this goal.

The field of Quality-control may be likened to the field of Industrial Safety in that both are based on the simple philosophy that an ounce of prevention is worth a pound of cure, that the dollar cost of preventing an accident to men or materials is considerably less than the cost of the cure. We can look back with a great deal of admiration on the work of the pioneers in the field of Industrial Safety who labored long and hard to get their methods accepted into industry. Proving to management that industrial safety is a sound dollar investment was a long uphill road, and the reason for this is that the benefits of a safety program do not appear in the form of immediate dollar savings but are instead in the nature of ling-term gains of intangible value. We are most fortunate that this is not the case with Quality-control, rather the first step in the presentation of a QC program to management is the computation of the exact amount of "gold in the mine" and an accurate forecast of the actual dollar savings which may be expected from the adoption of QC methods. This has been one of the most important factors which have resulted in such a rapid development in such a relatively short period.

Quality has long been an important part of the competitive picture in all industrial countries. Brand names are the direct result of the belief on the part of the buying public that the product bearing the name of a particular manufacturer will consistently meet the expectations of the buyer. The supply and demand mechanism of the market is constantly working to improve quality and at the same time to lower the price of goods. If a Quality-control program permits one manufacturer to sell better goods at lower prices, his competitors must adopt similar methods or suffer the pinch where it hurts the most--in the profit margins.

As a result of increased emphasis on quality, it has become an obligation on the part of top management to be quality-minded. It is the attitudes and interests of top management which condition the behavior of the operator on the plant floor. To perform their jobs well intermediate executives must reflect the interests of the men at the top and these interests are thus passed on all the way down the chain to the men who run the machines and the men who inspect the output of those machines. Unless continuing emphasis is placed on quality by top men, quality is bound to deteriorate.

The success or failure of any quality-control program may well hinge on the degree of quality-mindedness which is instilled in the employees. It is this quality-mindedness which results in all members of the industrial team pulling together to achieve better production at lower costs. Management should make use of as many as possible of the available media for selling quality to the operators: posters, films, competitions, and awards, and any other device which will stress to the employee that the management of the company is proud that the quality of its product is high. This is especially important as it pertains to new employees because, whether or not they are usually correct, first impressions are certainly difficult to counteract. Too few industries take advantage of the literature which is given to new employees to stress the company's interest in high quality output. If many companies would divert ten percent of the money they now put into advertising quality to the consumer of the product into selling quality to the producer of the product I'm sure that the results would be a pleasant surprise.

While forecasts of immediate dollar savings have greatly assisted in furthering the acceptance of Quality-control methods by management, we

have learned from experience that there are many other benefits of an intangible nature which derive from a QC program as by-products.

The benefits resulting from quality-mindedness in employees are not soley concerned with quality of output, they are also closely tied in with employee morale. Back in the early part of the century a great American social scientist, Thorstein Veblen, phrased what he called the "instinct of workmanship" in man, the desire of normal human beings to engage in constructive pastimes and their pride in a job well done. It is to this pride of workmanship that a properly conducted program of quality-control relates. It is unfortunate to have to realize that in some cases labor unions in their desire to further the cause of the working man have placed so much emphasis on monetary values that the worker's instinctive pride in his work is completely overshadowed. Quality-control, properly administered, can go far toward correcting or counteracting such tendencies. Generally speaking, HIGHER QUALITY MEANS HIGHER MORALE.

The question of employee morale is especially important in conditions such as exist today when employment is exceptionally high. Full employment means competition for labor and those companies which can boast high morale benefit in two directions. First, healthy employee morale will tend to discourage men presently employed from seeking work elsewhere, either in or outside the company, and second, new help will be more readily forthcoming from the labor market if company morale is high.

The recently developed and fast-growing field of Human Relations or Human Engineering is an outgrowth of the recognition of other-than-monetary motives in industrial employees. A number of surveys have disclosed that in most cases money is not the foremost concern of the worker, often ranks only sixth or seventh on the list of what the worker wants from his job. Usually higher on the list is the desire for recognition of work well done and the desire on the part of the worker to feel that what he is doing is important to the process of which he is a part. Many of our more progressive companies have made and continue to make large investments in satisfying these desires, spending a great deal of time and money in educating the employee to the part his job plays in the overall picture and giving him thereby the knowledge that he is an integral part of the industrial process rather than a neglected and unimportant detail. Chart Control procedures can go far toward furthering this education. The emphasis on good workmanship which these and other quality-control techniques place before the worker tend to make him feel that his job is appreciated and important. Chart Control supplies a measure of performance by which the worker can evaluate himself and in so doing the worker gives added interest to his job regardless of how routine and dull it may be. I remember walking along a lonesome country road as a boy and passing the time by counting the number of consecutive steps I dared take with my eyes closed, then trying to better my performance. I'm sure that each of you can recall similar situations, simple devices we used to add interest to an otherwise monotonous pastime. Chart Control supplies the worker with just such a simple device.

At the same time Chart Control gives supervisory personnel an exact measurement of operator quality performance, replacing the method of personal opinion for purposes of merit rating and advancement. Here again is another instance of substituting an exact statistical method for the exercise of human discretion.

A poll of operators in one plant taken after the introduction of a quality-control system revealed the following interesting majority opinions:

1) That the operator's understanding of the job was increased,
2) That more responsibility for product quality was placed on the operator,
3) That the operator was better able to control quality,
4) That operator-inspector relationships were improved.

These things are merely opinions of the particular persons involved, they do not stand as conclusive evidence that a quality-control program will achieve similar results, they do not even conclusively prove the success of the particular system in question. What they do prove is that a healthier attitude on the part of the employees taking part has resulted.

The point that operator-inspector relationships were improved is one of major importance. Where the operator feels that the inspector's job is to reject as much of his work as possible a serious hindrance to smooth output develops. Too often the inspector's methods remain a mystery to the operator who is forced to accept the inspector's decision without question. The quality-control program replaces the inspector's function as the administrator of acceptance standards with completely impersonal quality-control tools which may be administered by the operator himself, thereby eliminating the possibility that friction may arise. The operator becomes equally as able as the inspector to point out to himself whether a piece of work is or is not acceptable; the quality criteria reject below-standard work, not the inspector.

In putting an informal system of inspection into the framwork of a quality-control program we may usually expect to achieve an ordered presentation of facts which will reveal many important things hitherto obscured in the mass of detail. Repeatedly we receive reports from companies using quality-control that a completely new knowledge of quality standards has resulted from putting the program into effect. In many cases quality may have been unnecessarily high, a condition which may be just as costly as when the reverse is true. From any Quality-control program worth its salt should come a substantially increased ability on the part of those in control to set standards of quality which are neither too low nor too high. You or I would not pay a penny more for a comb because we know that its teeth are accurately spaced to a millionth of an inch.

Closely allied with the ability to see quality standards in a clearer light is the increased knowledge of machine performance capabilities which results from proper methods of quality-control. Accurate determination of performance characteristics permits a more intelligent decision as to when machinery is ready for replacement, and conversely, when it is not. We are told by Mr. Charles D. Wiman, President of Deere & Company, that Control Charts prevented that company from the unnecessary expenditure of ninety thousand dollars for new machinery when the charts proved that the old machines could do the job.

Another by-product which we can expect to derive from a quality-control program is a systematized series of reports which will keep management informed on the status of quality and also on the progress of the quality-control program. These reports if soundly constructed will

assist in presenting the overall picture to all concerned and materially assist in training all personnel to think in terms of the overall functions and results. It is this type of thinking which we especially want to engender in personnel being trained for positions of high responsibility.

In a context of executive training quality-control can play an important part. Personnel administering quality-control programs are in a particularly good position for becoming familiar with many facts of corporate operation. The Sales Department of any company is vitally interested in quality-control because it is so closely interrelated with sales functions; the Purchasing Department must be concerned with extending quality-control standards to incoming material; the Production Department must maintain close contact to administer the various production control devices; the Maintenance Department must assist in deriving and using machine performance information; the Accounting Department must accumulate all cost information; and so on down the list. Personnel, Engineering, Manufacturing, Standards--all have a vital interest in the functioning of the quality-control program. As a result of this, quality-control personnel cannot help but become familiar with many operations outside their immediate jurisdiction. Quality-control is so closely interwoven with so many activities that it is sure to afford an excellent education opportunity for those involved in it.

Certainly one of the most important criteria of the success of a QC program is the degree to which materials for scrap and rework are cut back. The dollar savings of such a cut-back as they pertain to the cost of materials, wasted labor and labor of reworking rejected materials are obvious. But there are other savings which we are apt to overlook in computing the value of QC results. The paperwork connected with rework labor and labor wasted on materials later scrapped has always been a hindrance to the smooth functioning of the production control system, and costs of such additional operations must be carried through to the accounting department for costing and payroll purposes; decrease in the number of rejections means decrease in the costs of these functions. Rejection of materials also usually means additional burdens on the internal transportion system, and added wear on machines assigned to rework operations. Improvement in these conditions means additional savings, lower production costs.

I would like to inject here a word of caution regarding the installation of a quality-control program. Too many times failure has resulted from trying to accomplish too much in a short time. Unchecked enthusiasm on the part of the person or persons administering the QC program may mean the difference between success and failure of the entire operation. Remember that to be successful any program must be thoroughly understood by the men who will handle the day-to-day details of it. There are two difficult selling jobs connected with the installation of a QC program: First, to sell it to management, and second, to sell it to the operator. The importance of this second point cannot be overstressed. Remember that you are dealing with statistical tools which, at first glance, are as incomprehensible to the average workman as the Chinese language; to get these tools into practical everyday use they must be presented in a manner so simple as to permit complete comprehension by the persons who will make daily use of them. And such a presentation is bound to be a slow and difficult process. Unless you do a successful job of instilling your comprehension and enthusiasm into the operator all your efforts are

doomed to failure. It is human nature to take a firm stand against anything we do not understand. Knowing this we have no right to expect acceptance of quality-control by anyone to whom its operating principles are not clear. The amount of time you spend working out simple presentations of basic operating principles and in constructing simple devices to demonstrate these principles will pay ample dividends in smoothing out the rough spots which any new system is bound to encounter.

If operator enthusiasm is properly cultivated you will find that it can become a force with the strength of a tidal wave. Many companies have found that it is next to impossible to discontinue the use of some quality-control devices because the operators in the plant will not stand for it.

In conclusion I would like to restate briefly what I mentioned at the outset of this talk, that I feel that the industrial world is on the verge of a completely new frontier of scientific application and advance, and that we have every reason to expect that American industry will pioneer in the exploration of that frontier. By way of contrast, I am reminded of that well-known worrier of the 18th century in England, Mr. Thomas Malthus, with his theory that the population was growing at such a rate that it was certain to soon outstrip the supply of food and result in starvation for great hordes of people. This type of pessimism was not confined to early England; we too have our modern prophets of doom. Today they are concerned with the exhaustion of natural resources and the industrial and economic stagnation of the American economy. They ignore the fact that the very concept of what makes up natural resources is a dynamic thing, that a mineral which is vital to our lives today may be of relative unimportance tomorrow. The gasoline which now propels our vehicles may give way tomorrow to a completely different kind of fuel, experiments in solar heating may make the coal or oil burner obsolete for heating purposes, plastics continue to replace metals in many applications, and synthetic cloth gains in the competition with natural fibres. There are many such examples of similar changes which are taking place around us at this very moment.

Nor can I find anything constructive in this attitude that we should fear what the future holds in store. The reason that problems like that of the exhaustion of natural resources are solved is that men have enough faith in the future to work calmly and steadily to better the present. As long as this remains true the present remains a dynamic and progressive force; with every successful experiment something out of the future becomes a part of the present.

We are part of one such experiment, part of one such success. The rapid rise of Quality-control to its present position was in itself the solution of a problem, the problem of systematizing the industrial framework of the present to accomodate the methods of the future.

LINEAR PROGRAMMING

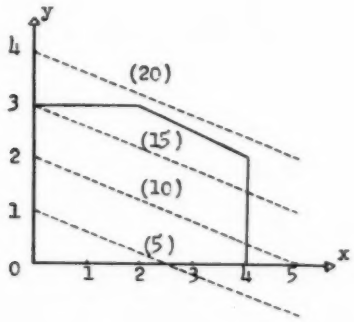
A. W. Tucker
Princeton University

Linear Programming deals with maximizing or minimizing a linear function of several real variables subject usually to the requirement that the variables be nonnegative and satisfy certain additional linear inequalities (or equations). The following problem is an illustration, in miniature.

EXAMPLE 1. To maximize $2x + 5y$ subject to the five constraints

$$x \geq 0,\ y \geq 0$$
$$x \leq 4,\ x + 2y \leq 8,\ y \leq 3\ .$$

This problem can easily be solved graphically since just two variables occur in it. The set of feasible points (x, y) satisfying all five constraints is a pentagonal region bounded by the y-axis $(x = 0)$, the x-axis $(y = 0)$, and the lines $x = 4$, $x + 2y = 8$, $y = 3$, as shown.



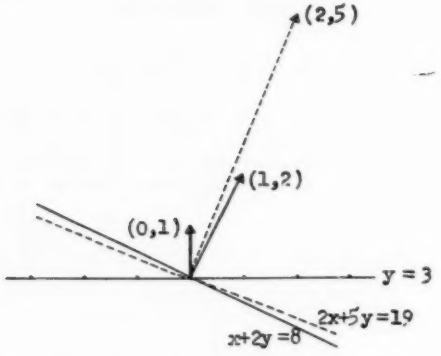
The function $2x + 5y$, which is to be maximized, is represented by contour lines (broken) along which it takes the values 5, 10, 15, 20, as indicated. We see that the desired maximum seems to occur at the corner determined by the intersection of the two constraint lines

$$x + 2y = 8,\ y = 3\ ,$$

namely, at the point $x = 2$, $y = 3$, for which

$$2x + 5y = 19\ .$$

In order that this be the desired maximum it is necessary that the contour line $2x + 5y = 19$ should not enter into the interior of the pentagonal region of feasible points.



Hence the vector $(2, 5)$ normal to the contour line $2x + 5y = 19$ must belong to the "angle" spanned by the vectors $(1, 2)$ and $(0, 1)$ normal to the constraint lines $x + 2y = 8$ and $y = 3$, respectively. That is, there must exist nonnegative coefficients v and w such that

$$(2, 5) = v(1, 2) + w(0, 1) ,$$

which requires

$$v = 2, \ 2v + w = 5 .$$

The nonnegative solution $v = 2, w = 1$ of this pair of auxiliary equations yields the breakdown

$$2x + 5y - 19 = 2(x + 2y - 8) + (y - 3) .$$

Hence $2x + 5y \leq 19$ for all (x, y) such that $x + 2y \leq 8, y \leq 3$. This provides an algebraic check that the feasible point $x = 2, y = 3$ gives $2x + 5y$ its maximum value 19 subject to the five stated constraints.

EXAMPLE 2. To <u>minimize</u> $4u + 8v + 3w$ subject to the five constraints

$$u + v \geq 2, \ 2v + w \geq 5 ,$$
$$u \geq 0, \ v \geq 0, \ w \geq 0 .$$

Here the set of feasible points (u, v, w) satisfying all five constraints is a pentahedral region bounded by the planes $u = 0$, $v = 0$, $w = 0$, $u + v = 2$, $2v + w = 5$. It consists of the positive octant $(u \geq 0, v \geq 0, w \geq 0)$ with two prisms removed, the one bounded by the three planes $u + v = 2$, $u = 0$, $v = 0$ and the other by the three planes $2v + w = 5$, $v = 0$, $w = 0$. It possesses three corner, or extreme, points:

$(0,\ 2\ ,\ 1)$, the intersection of $u + v = 2,\ 2v + w = 5,\ u = 0$,
$(2,\ 0\ ,\ 5)$, the intersection of $u + v = 2,\ 2v + w = 5,\ v = 0$,
$(0,\ 2.5,\ 0)$, the intersection of $2v + w = 5,\ u = 0,\ w = 0$.

At these three extreme feasible points the function $4u + 8v + 3w$, which is to be minimized, takes the values 19, 23, 20, respectively. To make sure that the point $(0, 2, 1)$ minimizes $4u + 8v + 3w$ subject to the five constraints, it is necessary to express the vector $(4, 8, 3)$ normal to the contour plane $4u + 8v + 3w = 19$ as a nonnegative linear combination of the vectors $(1, 1, 0)$, $(0, 2, 1)$, $(1, 0, 0)$ normal to the constraint planes $u + v = 2$, $2v + w = 5$, $u = 0$, respectively:

$$(4, 8, 3) = x(1, 1, 0) + y(0, 2, 1) + z(1, 0, 0) .$$

That is, it is necessary to solve

$$\begin{gathered} x \geq 0,\ y \geq 0,\ z \geq 0 , \\ x + z = 4,\ x + 2y = 8,\ y = 3 , \end{gathered}$$

or, setting $z = 4 - x$, to solve

$$\begin{gathered} x \geq 0,\ y \geq 0 , \\ x \leq 4,\ x + 2y = 8,\ y = 3 . \end{gathered}$$

The solution $x = 2$, $y = 3$, with $z = 2$, yields the breakdown

$$4u + 8v + 3w - 19 = 2(u + v - 2) + 3(2v + w - 5) + 2u .$$

Hence $4u + 8v + 3w \geq 19$ for all (u, v, w) such that $u + v \geq 2$, $2v + w \geq 5$, $u \geq 0$. This provides an algebraic check that the feasible point $(0, 2, 1)$ gives $4u + 8v + 3w$ its minimum value 19 subject to the five stated constraints.

By now it should be evident that Examples 1 and 2 are closely related. They involve the same data and have a common optimum value, 19. Moreover, a check of the solution of either one rests on a breakdown in terms of the solution of the other. The common structure of the two problems is exhibited conveniently by the following scheme of "detached coefficients."

(≥ 0)	x	y	$(\leq)$
u	1	0	4
v	1	2	8
w	0	1	3
$(\geq)$	2	5	

Example 1 pertains to the inner (dot or scalar) products of rows in this scheme with the initial row (x, y); we seek a point or vector (x, y) that maximizes the inner product of $(2, 5)$ with (x, y), namely $2x + 5y$, subject to the individual components of (x, y) being nonnegative and to the inner products of the rows $(1, 0)$, $(1, 2)$ and $(0, 1)$ with (x, y) not exceeding 4, 8 and 3, respectively. Example 2 pertains to inner products of columns in the scheme with the initial column (u, v, w); we seek a point or vector (u, v, w) that minimizes the inner product of $(4, 8, 3)$ with (u, v, w), namely

$4u + 8v + 3w$, subject to the inner products of the columns $(1, 1, 0)$ and $(0, 2, 1)$ with (u, v, w) being not less than 2 and 5, respectively, and to the individual components of (u, v, w) being nonnegative.

The solutions already obtained can be readily checked by substituting in the scheme.

(≥ 0)	2	3	$(\leq)$
0	1	0	4
2	1	2	8
1	0	1	3
$(\geq)$	2	5	19

The final step is to test the two solutions against one another by forming the inner product of $(2, 5)$ with $(2, 3)$ and of $(4, 8, 3)$ with $(0, 2, 1)$. Both inner products are 19, the common maximum and minimum. This completes the check.

A general pair of dual linear programming problems is presented by a similar matrix scheme.

(≥ 0)	x_1	x_2	$\cdots$	x_n	$(\leq)$
u_1	a_{11}	a_{12}	$\cdots$	a_{1n}	b_1
u_2	a_{21}	a_{22}	$\cdots$	a_{2n}	b_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
u_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}	b_m
$(\geq)$	c_1	c_2	$\cdots$	c_n	

Here the "row" problem is to find a vector $X = (x_1, x_2, \ldots, x_n)$ that maximizes the inner product

$$C \cdot X = c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

subject to the m row constraints

$$(1) \qquad a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \leq b_i \qquad (i=1,2,\ldots,m)$$

and the n nonnegative constraints on the individual components of X

$$(2) \qquad x_j \geq 0 \qquad (j=1,2,\ldots,n).$$

The "column" problem is to find a vector $U = (u_1, u_2, \ldots, u_m)$ that minimizes the inner product

$$B \cdot U = b_1u_1 + b_2u_2 + \ldots + b_mu_m$$

subject to the n column constraints

$$(3) \qquad a_{1j}u_1 + a_{2j}u_2 + \ldots + a_{mj}u_m \geq c_j \qquad (j=1,2,\ldots,n)$$

and the m nonnegative constraints on the individual components of U

$$u_i \geq 0 \qquad (i=1,2,\ldots,m). \tag{4}$$

A vector X or U that satisfies the $m+n$ constraints (1) and (2) or (3) and (4) is called feasible. The test whether a feasible X or U yields the desired maximum or minimum is provided by the following "fundamental theorem."

THEOREM 1. A feasible X or U is a solution of the stated maximum ("row") or minimum ("column") problem if, and only if, there exists a feasible U or X such that $B \cdot U = C \cdot X$.

Since the rôles of the solution vector X or U and the "test" vector U or X in this Theorem are interchangeable, it is seen that the "test" vector is actually a solution for the minimum or maximum problem, respectively. That is, either both problems possess solutions or neither possesses a solution. The condition for solutions to exist is stated in the following theorem.

THEOREM 2. Solutions X and U exist if the two dual problems are both feasible, i.e. if both possess feasible vectors.

Proofs of these theorems rest on a basic theorem of J. Farkas [Journal für die reine und angewandte Mathematik, 124 (1902) pp. 5-7] which states that a homogeneous linear inequality (written ≥ 0) is a nonnegative linear combination of a system of such inequalities (all written ≥ 0) if it is satisfied by all solutions of the system. For details see the paper on "Linear Programming and the Theory of Games" by D. Gale, H. W. Kuhn, and A. W. Tucker, pp. 317-329 in Reference [1], below.

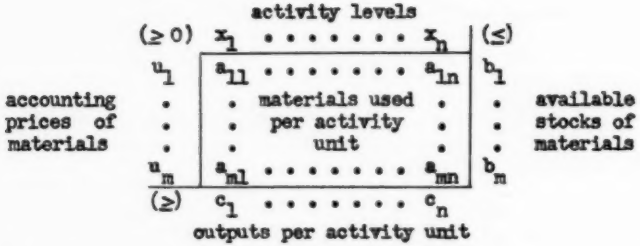
As an application of Linear Programming consider the following Activity Analysis problem (see Reference [1], below). Let there be n ways ("activities") of making a single desired commodity from available stocks of m primary commodities ("materials"). Let

$$a_{ij},\ b_i,\ c_j,\ \text{and}\ x_j$$

be, respectively, (a) the amount of the ith material used in one unit of the jth activity, (b) the available stock of the ith material, (c) the quantity of the desired commodity made by one unit of the jth activity, and (x) the number of units of the jth activity to be undertaken. Then, assuming "constant returns to scale," the problem is to find an "activity vector" $X = (x_1, x_2, \ldots, x_n)$ that will maximize the total output

$$C \cdot X = c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

subject to the m constraints (1), above, set by the available stocks of the m materials and to the n constraints (2), above, set by the natural impossibility of negative "activity levels."



The dual problem pertains to accounting (fictitious or "shadow") prices attached to the m materials, on a scale whose unit is the price of the desired commodity. One seeks a "price vector" $U=(u_1,u_2,\ldots,u_m)$ that minimizes the total accounting value

$$B \cdot U = b_1u_1 + b_2u_2 + \ldots + b_mu_m$$

of the available stocks of materials subject to the "one-can't-get-something-for-nothing" requirement (3), above, that the total accounting value of the materials used in one unit of each activity is not less than the accounting value of the quantity of the desired commodity produced by one unit of that activity, and the requirement (4), above, that the accounting prices be nonnegative.

The traditional calculus procedure for finding a constrained maximum or minimum employs Lagrange multipliers. These may be adapted to our linear programming problems by introducing the Lagrangian function

$$\begin{aligned}\phi(X,U) &= c_1x_1 + \ldots + c_nx_n + u_1(b_1 - \textstyle\sum a_{1j}x_j) + \ldots + u_m(b_m - \sum a_{mj}x_j) \\ &= b_1u_1 + \ldots + b_mu_m + x_1(c_1 - \textstyle\sum a_{i1}u_i) + \ldots + x_n(c_n - \sum a_{in}u_i).\end{aligned}$$

In the first expression on the right the u's are Lagrange multipliers applied to the maximum ("row") problem while in the second the x's are Lagrange multipliers applied to the minimum ("column") problem. This bilinear symmetry is very gratifying but, of course, just reflects the fundamental duality already observed.

THEOREM 3. X and U are solutions of the maximum ("row") and minimum ("column") problems if, and only if, they determine a "saddle-point" at which the Lagrangian function $\phi(X, U)$ has a minimax value, minimum as regards all nonnegative U and maximum as regards all nonnegative X.

This theorem is closely related to Theorem 1, above; a proof of a more general minimax theorem has been given by H. W. Kuhn and A. W. Tucker in "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, 1952) pp. 481-492. In the activity analysis problem, above, $\phi(X, U)$ is the combined accounting value of the total output $C \cdot X$ of the desired commodity and of the unused balances of the materials.

The computation of solutions of linear programming involving a large number of variables has been and continues to be an important research task of the Department of the Air Force and of the National Bureau of Standards (see Reference [2], below). Dr. G. B. Dantzig, an outstanding pioneer in the field, has devised an ingenious "simplex method" of solving linear programming problems or general systems of linear inequalities and equations, analogous to the Gauss elimination method for solving systems of linear equations (see Reference [1], pp. 339-373). It is generally true, as Examples 1 and 2 above indicated, that the corner or extreme solutions of linear programming problems are just the solutions of a subsystem of linear equations obtained from the subset of constraints that determine the corner or extreme feasible point. But, with a large number of variables, it becomes a very complex matter of trial and error to isolate the constraints that determine a corner or extreme solution.

Reference [3], below, provides an excellent introduction to linear programming, its methods and varied applications, growing out of work done at the Carnegie Institute of Technology for the Department of the Air Force.

Acknowledgment: This report is based on work done for the Logistics Project, Princeton University, and the Logistics Research Project, The George Washington University, under contracts with the Office of Naval Research.

References

[1] T. C. Koopmans (editor), ACTIVITY ANALYSIS OF PRODUCTION AND ALLOCATION (Cowles Commission Monograph 13), New York: John Wiley and Sons, 1951.

[2] A. Orden and L. Goldstein (editors), SYMPOSIUM ON LINEAR INEQUALITIES AND PROGRAMMING, Washington, D. C.: Project SCOOP, Comptroller, Headquarters U. S. Air Force, 1952.

[3] A. Charnes, W. W. Cooper, and A. Henderson, AN INTRODUCTION TO LINEAR PROGRAMMING, New York: John Wiley and Sons, 1953.

THE ROLE OF A BIOMETRICIAN IN THE PHARMACEUTICAL INDUSTRY

C. I. Bliss
The Connecticut Agricultural Experiment Station
and Yale University

The only word in my title that may be new to some of you is "biometrician", so perhaps I should begin with a definition. In general, a biometrician is a scientist concerned with the statistical and mathematical aspects of biology. In the development and manufacture of drugs -- defining the term broadly to include the vitamins, hormones, biologicals, antibiotics and related preparations -- he is concerned primarily with the design and evaluation of experiments with a living indicator, such as a bacterium, an animal, a plant, a bacteriophage, an isolated tissue or a human patient. Although the dose of drug may be known exactly from chemical or physical measurements, the response is subject to a type of variation which the biometrician is especially qualified to handle. In other experiments the reaction may not be measured biologically, but the biological origin of the drug often calls for a similar approach.

Thus his subject matter distinguishes him initially from statisticians who are concerned with the chemical, physical or engineering problems arising in the manufacturing process. Nor should he be confused with statisticians who handle questions of marketing and economic forecasting. From his duties, the biometrician belongs on the same team with the biochemist, immunologist, bacteriologist, clinical investigator and manager of biological control, in short, on the team of professional biological scientists upon which the industry depends. His contribution, however, differs from that of any of these, as will be evident from his qualifications.

Qualifications. We may consider the qualifications of a biometrician in the pharmaceutical industry in terms of his training and of the statistical and mathematical methods he should be able to use. Ideally, he should have a Ph.D. with a major in statistics and a minor in pharmacology or biology. Under statistics, he will need especially courses in experimental design and analysis and at least an introduction to mathematical statistics. His mathematical background should be as strong as possible, so that he can read and apply new techniques as they are developed. Moreover, biometry includes the mathematical as well as the statistical aspects of biology and these are bound to crop up in time. Biology, however, is almost equally important and for our present purpose should include pharmacology, bacteriology, physiology, biochemistry and the background sciences.

In the drug industry many of the statistical methods that a biometrician should be able to use effectively fall under the general heading of the statistics of bioassay. Although a development of the last 20 years, a considerable literature of papers and books has developed, ranging from slender volumes (7, 19) to the extended treatment in Finney's "Probit Analysis" (20) and "Statistical Methods of Biological Assay" (21). In applying these techniques, he will design experiments so as to

segregate extraneous sources of error. He will estimate the number of observations needed for a given precision. He will compute tests of significance and confidence limits under varied conditions. He will plan factorial experiments for studying several factors simultaneously. When the experimental results can be obtained serially, he will proceed sequentially, so as to minimize the number of observations. In applying these methods, he will use the analysis of variance, regression, interval estimation, covariance, variance components and other statistical tools. Applications of many of these may be found in the proceedings of a conference on the place of statistical methods in biological and chemical experimentation (34). Here they will be considered in relation to the stages a new drug must traverse from its birth until it is ready for the market, beginning with research and development.

Screening programs for new drugs. The research laboratories of a drug house are constantly seeking new drugs, whether synthetic or of natural origin. As each compound is prepared, its biological activity must be tested. Pharmacological screening can be expensive and may take many forms. Prominent among these are tests involving an all-or-none response. For this type of screening, the appropriate design will depend upon how promptly the animal reacts after it is given the drug. When the reaction time is short, the biometrician may recommend a staircase or up-and-down design. With this technique, the response of one animal to its dose determines the dose to be given to the next animal. If the reaction is positive, the dose is reduced; if negative, the dose is increased, in each case by a constant log-dose step (20). If, on the other hand, the reaction time is relatively long, a small but constant number of animals, from two to six or ten, may be assigned at random to each of four dosage levels, spaced at equal log-intervals. Tables have been published recently with which the ED50 and its standard error for this particular design can be estimated with a minimum of calculation (35).

The precision sought in each test will depend in part upon the probable cost of producing the drug and on the risk assumed in missing a potentially valuable product. With an all-or-none reaction precision can be increased both by adding more animals and by adopting a more penetrating method of statistical analysis (3, 20), one which will provide a measure of the slope and tests for homogeneity and parallelism. Similar differences in the intensity of screening are available with graded or measured reactions. To recommend the most efficient design for each purpose is one of the many functions of the biometrician.

The pharmacological characterization of a new compound. A compound that passes the initial screening is then subjected to a battery of pharmacological tests, many of them primarily quantitative. Qualitative experiments, such as for side effects, are usually delayed until the activity of the compound is known. Hence, the first tests are likely to be quantitative determinations of potency. Potency can seldom be expressed satisfactorily in terms of a biological response; the response per se is subject to too many environmental and inherent biases. These are avoided by comparing the reaction to a new drug with the concurrent re-

sponse to some other drug. Hence an estimate of potency is essentially a comparative bioassay and should be designed statistically and carried out as such. Unless the therapeutic action of the compound is unique, a standard drug in general use serves as the reference standard. In effect it calibrates the assay organisms in respect to their sensitivity.

A drug intended for use against a pathogen, such as one of the antibiotics or antimalarials, is assayed initially in vitro with one or more strains of microorganisms and only later in animals inoculated experimentally with the disease. A much higher level of precision is to be expected, of course, in the first type of experiment. In penicillin, for example, the standare error of an individual cylinder plate assay may be less than the laboratory errors in handling the drug (5, 11). The situation is quite different, however, when the drug is tested against pathogens in inoculated animals. In tests of antimalarials in ducklings, for example, the percentage of parasitized red blood cells varied far more than could be accounted by the usual sampling error for an all-or-none response (32).

With other classes of drug, the pharmacological activity is measured directly. Many examples have shown how the biometrician can increase the efficiency of a pharmacological test. Experiments on vitamin K, for example, have been based upon the clotting time of chick blood. Some investigators thought they could improve upon a published technique by bleeding the chicks twice, before as well as after administering the vitamin K. They reasoned that knowing the initial blood-clotting level of each chick would lead to a more precise estimate of the effect produced by vitamin K. Later statistical analysis proved that just the opposite was true (10). Here the trauma from the initial bleeding actually reduced the inherent precision of the assay. Or again, a covariance study of the parathyroid assay showed that the initial blood level could be dispensed with in a suitably designed experiment, thus cutting the number of chemical analyses in half (13). Experimental and biometric studies of the insulin assay initiated by the USP Insulin Advisory Board give promise of reducing the number of blood sugar determinations even more drastically without sacrificing precision. No doubt, many other tests for pharmacological activity could be improved equally by a quantitative study of sources of error, arbitrary assumptions, and laboratory routines, often by simple rearrangements of routine experiments without recourse to additional tests not now provided in the budget.

Before a new drug can be released, its toxicity in the host as well as its potency must be determined. A quantitative estimation of relative toxicity (3, 20) in one species of laboratory animal is not sufficient, however, when the inference is to man. The toxicity of the drug should be assayed instead on several different species. Relative toxicity may differ markedly from one laboratory animal to another, as was observed with certain alkaloids (15). When this occurs, the new drug's toxicity in man is much more doubtful than if its toxicity relative to a suitable reference standard were the same in all species. Moreover, the new drug is usually tested on healthy laboratory stock although it is to be used in sick patients, adding an additional hazard

to extrapolation to man.

Of particular importance is the chemotherapeutic index of a new drug or the ratio of its toxic to its therapeutic dose. When the dose effecting a 50% cure is compared with that for a 50% kill, the index may be a poor guide, for then it ignores the effect of the slope of the dosage-effect and dosage-mortality curves. A very small slope in one or both of these curves would disqualify a prospective drug. Instead, the chemotherapeutic index should be computed from the dosage curing a high percentage, such as 95 or 99%, and killing a small percentage, such as 1 or 5% (22). The 5 and 95% levels can be interpolated from the curves, with smaller sampling errors than the 1 and 99% levels. Both tend to be less stable than the 50% levels, but this limitation can be offset in part and the sampling error of the index reduced by the application of statistical criteria in selecting dosage levels and in computing the results.

Clinical trials. After a new drug has performed satisfactorily in animal tests, it is ready for clinical trial. Sometimes samples of a new drug are distributed to selected physicians, who are left to their own devices as to how they shall conduct their clinical trials. Most physicians are ill-equipped either to design critical experiments in this field or to interpret their data. The result may be a series of conflicting reports from which the truth emerges only after a considerable period. Many difficulties could be avoided if all clinical trials were designed and analyzed with the aid of a good biometrician. Some drugs, of course, are so outstandingly effective that the correct answer sticks out like a sore thumb. More often, however, the improvement is less marked and can only be evaluated by careful statistical analysis. There are currently so few clinics that boast the services of a competent statistician or biometrician that in most cases both the drug and the biometrical advice will have to come from the manufacturer if reliable results are to be assured.

In many clinical trials it is only possible to state that one drug has been more effective than the other in a comparative trial, one of which may be a placebo. In this type of experiment, the necessity of concealing the identity of the treatment not only from the patient but also from the attending physicians and nurses, the so-called "double blind" test, cannot be stressed too strongly. The danger of biassing the outcome in experiments on analgesics and hypnotics has been discussed recently in Science (1, 2), but it is also important in other areas where the psychological factor may be overlooked, as in testing the effect of anticoagulants upon the number of coronary thromboses and even on the number of fatal outcomes (23). Without sound original data, no amount of statistical analysis can lead to a valid conclusion.

The experimental results of a clinical trial can often be summarized in a 2 × 2 table. These may show the number of patients cured and the number not cured on each of two alternative treatments or may take several other forms (8). The analysis differs with the design but leads in each case to an estimate of the probability that the observed results might

have occurred if there were no inherent difference between the two treatments. Mosteller (28) has recently noted some of the precautions needed in tests of this type with analgesics.

When many factors are involved, the data cannot be summarized in a single 2 × 2 table. Instead we have a complex contingency table with computational difficulties at the professional level. This occurred, for example, in a comparison of ether and cyclopropane anaesthesia in relation to the incidence of circulatory complications after abdominal operations (18). The outcome depended upon a number of conditions, such as the site of the operation (upper or lower abdominal), status of the circulatory system (normal or abnormal), and degree of complication (major or minor). A tendency to use each anaesthetic in a particular type of case could bias the conclusions if the different factors were not isolated.

When the results can be determined only after a long delay, the number of cases needed for discriminating between treatments has to be estimated in advance. In other trials, however, each result can be assessed relatively soon and then the number of observations can be determined by the outcome. A practicable sequential program has been described recently by Bross (14). Cases are paired as they appear, assigning the patients in each pair to treatments at random. If both individuals in a pair react alike, either positively or negatively, that pair cannot contribute to discriminating between the two treatments. Discrimination depends on the pairs in which one patient reacts differently from the other. The physician makes four decisions in advance, probably with the aid of a biometrician: (1) how much better must the new remedy be before he would consider it an improvement, (2) what is the probable rate of cure with the standard or control preparation, (3) what protection does he want against concluding that there is a difference when none exists (against making an error of the first kind), and (4) what risk will he assume of missing a real difference by calling the two drugs alike when in fact they differ (an error of the second kind). Given these decisions, a testing scheme can be devised which requires no computation, not even addition, and leads with a minimum number of trials to one of three conclusions: (1) the new drug is better than the old drug, (2) the two do not differ significantly, or (3) the old drug is better than the new drug.

Another danger in clinical experimentation is that the results may be obscured by individuals whose reaction is irrelevant, as shown by Jellinek's study of two ingredients in a compound headache remedy (26). Four identically-appearing kinds of tablets were prepared, one being the complete drug, the second had one ingredient missing, the third lacked a different ingredient and the fourth was a placebo. As patients with chronic headaches came to see their physicians, each was given a package of tablets, to be taken every time he had a headache. He was to note whether or not his headache was relieved within half-an-hour. At the end of two weeks, each subject reported how many headaches he had had and how many of these had been relieved satisfactorily. He was then given a fresh envelope containing one of the other drugs and this was repeated until each patient had tried all four drugs. Neither the patient nor the physician knew

the composition of the tablets, which were so distributed that all preparations occurred in the first, second, third and fourth two-week period as determined by a 4 × 4 Latin square.

An analysis in terms of the success rate showed no significant differences among the four drugs. However, it seemed probable that patients reacting positively to the placebo had psychological rather than physiological headaches and, therefore, were not suitable test subjects. The data were then reanalyzed, separating the 120 subjects who reported relief with the placebo from the 79 who found it ineffective. In the first group the success rates were almost identical for the other three drugs, varying from 82 to 87%. Among the subjects not reacting to the placebo, the success rate varied from 67 to 88%, a significant difference. By omitting the unsuitable subjects, the true clinical effectiveness of the ingredients could be evaluated, the highest rate occurring with the complete drug.

When there are more preparations to test than can be given to a single patient, incomplete balanced blocks offer a solution. This design was used in a comparison for intramuscular pain of six samples of penicillin sodium which differed in their content of impurities (25). Each patient was injected with three of the preparations at 3-hour intervals, so that every sample of drug occurred equally often with every other sample. Intensity of the pain following each injection was scored from 0 to 4 by comparison with a symmetrical injection of saline alone. Thus differences between patients could be segregated and, by a factorial plan, the sensitivity of site of injection tested simultaneously.

In other clinical trials, the magnitude of the response may depend upon external factors which could modify the conclusions rather importantly but which require statistical analysis for their discovery and measurement. In experiments at the Cornell University Medical College, the dose of digitoxin which just produced a recognizable change in the RT-T segment of the electrocardiogram was determined in each of 89 patients, ranging in age from three to 74 years. An initial report (27) indicated that 50% more digitoxin was required in children than in adults to produce a just detectable response, but this apparent contrast between children and adults vanished when the same observations were reanalyzed by partial regression and covariance (9). Here the age effect was due solely to an incorrect adjustment for differences in size. In the first report the dose per patient was divided traditionally by his body weight; in the later analysis by the square root of body weight as determined by statistical analysis. The relation of dosage to body size may differ with the type of drug but its correct evaluation in each case might resolve perplexing differences between competing drugs having the same type of action.

Many of the basic principles of biological assay are as applicable in the clinic as in the laboratory. In a recent study on the potency of mercurial diuretics with patients in congestive failure (24), the experiment was arranged in randomized groups, each group representing the reactions of a single patient to two or three different doses of each drug, given in a random order. The weight loss in quarter-pounds 24 hours after each

administration of drug served as the response. Since known differences in the dosage of each drug produced highly significant differences in response, bioassay procedures could be used to measure how much of each new drug was required to produce the same weight loss as a specified dose of the standard drug. When clinical experiments of this type can be carried out effectively, they are perhaps the most informative of all in assessing the relative potency of a new preparation.

Some topics in pharmacological research. Many problems arise in the pharmacological research laboratory of a manufacturer, where aid from a biometrician could be invaluable. One of these concerns the presence of synergistic or antagonistic action in a drug combination. Drug combinations are sometimes patented and they must then be shown to have an enhanced effect over what would be expected if each component were to act separately (4, 30, 31). This requires setting up suitable mathematical models and testing the response observed with a mixture against that predicted from the reaction to the same drugs when they are applied separately. Unless the mixture shows a significantly greater (or, in the case of antagonism, smaller) effect than can be predicted from their action alone, no claim of synergism (or antagonism) can be sustained. Those of you who have read the recent contributions to this subject need no convincing that here is an area where the services of a good biometrician are indispensable.

In the search for new drugs, series of related compounds are often evaluated quantitatively, even though they offer no early prospect of clinical value. Such observations may lead to a better understanding of the mode of action in relation to chemical structure, which in turn can guide the research chemist. When these compounds are tested over a period of years, comparison of results obtained at different times is open to question. A case in point is the relative potency of α-antiarin and β-antiarin in cats (16). These isomers seemed to differ significantly in terms of their LD50's when tested six months apart (26.2 ± 7.2), but the difference vanished when they were retested concurrently (3.1 ± 6.7). The initial significant difference was apparently due to unequal sensitivity in the test animals on the two occasions.

Bioassays have a theoretical appeal to the biometrician. Ordinarily when he draws inferences concerning an underlying population and computes confidence limits measuring the range of his uncertainty, he cannot check his procedures objectively. This limitation does not exist in bioassays of known but undisclosed concentrations of drugs, where he can determine experimentally whether his estimated confidence limits are operating as he would predict from his mathematical models. This same principle of known unknowns has also been introduced into many collaborative experiments for validating a method, such as those sponsored by the U. S. Pharmacopoeia. Here the services of the biometrician have been called upon repeatedly.

The improvement of the production process. Another area where a biometrician may be extremely valuable is in the improvement of a production process. These may be pilot plant studies

or conducted with full scale equipment. Where the product depends upon the growth of an organism, as in the production of penicillin and other antibiotics, the biometrician has important assets. He will be familiar with devices such as sequential experimentation and fractional factorial designs. How the latter may be applied to improving the yield of penicillin has been described in two recent papers (17, 34), where a relatively small number of experiments led to marked gains in yield (7).

Biological standardization. A number of important drugs must be standardized biologically in production, although as our knowledge increases, bioassays are often replaced by physical and chemical techniques. This is a function of the control rather than the research laboratory. In each case the active constituent in a batch of drug is assumed to be chemically identical with the reference standard, but present at an unknown concentration in an inert carrier. The objective of the assay is to determine its relative potency for use in adjusting the concentration or the labelled claim of the finished product. This implies an analytical bioassay, which is to be differentiated from the comparative bioassays discussed previously. In analytical assays, the same relative potency should be obtained for a given new batch of drug no matter what the species of animal, the route of administration or the environmental conditions.

The assumptions in an analytical assay do not hold in most comparative assays, where we know in advance that the unknown and the reference standard differ qualitatively, or, for that matter, in many of the assays described in the Pharmacopoeia, which ideally should be analytical. On the contrary, in most official assays almost every detail is specified. In addition, each individual assay is self contained. Frequently, however, the assay procedure itself could be placed under quality control within a given laboratory. This not only would provide a continuing check upon the testing procedure but would also increase the precision of the estimated potency, within the framework of USP requirements. Let us see how this might operate.

Many routine determinations of potency depend upon a graded response by the test organism. Two or more dosage levels are administered of the standard, or of both the standard and the test preparation or unknown, from which the relation between the measured response and the log-dose of drug can be determined. This relation must be known before the difference in the mean response to the two preparations can be converted to dosage units. The assay process is characterized by three basic statistics. The first of these is the slope (b) of the dosage-response curve with the dose measured on a logarithmic scale and the response on a scale that has been shown to give a linear relation in developing the assay technique. The standard deviation (s) is the second statistic, computed from the variation in the response; it is essential for estimating the confidence limits of the assayed potency. The precision of the assay depends upon the ratio of these two statistics, or upon $\lambda = s/b$. Finally, when the dosage-response curve is determined for both preparations, their parallelism can be tested statistically. Divergence in slope may be due to a poor experiment or to a qualitative difference between the two preparations. In either

case it is a warning sign that usually calls for a repeat assay.

These key statistics are determined repeatedly as an integral part of each assay, but not infrequently the slope and standard deviation reveal a very useful stability when plotted on separate quality control charts. When the slope remains stable over a series of comparable assays, its combined value is determined with increasing precision. Soon the sampling error of the combined slope contributes negligibly to the confidence limits of potency. These may then be computed with a relatively simple equation. Similarly, the standard deviation may show no significant change for months or longer, due presumably to stability in the random factors upon which it depends. Especially when it is determined repeatedly under identical conditions, the so-called inefficient estimates from the range or modified range are adequate and much easier to compute. Occasionally, both the standard deviation and the slope are unpredictable but their ratio ($\lambda = s/b$) is stable. Although less satisfactory than when the components are constant, a λ in control can still better the precision of an assay (7).

In most assays the standard deviation of a graded response is not only assumed to be stable from one dosage level to another but also to be distributed normally. This was tested quantitatively in a recent study (12) of the liver glycogen assay of adrenal cortex extract. Adrenalectomized rats were assigned at random in groups of five to the different dosage levels and preparations. From the known distribution of the range in samples of five from a normal population, the observed variation in response could be tested for agreement with the normal curve, even though the sample means differed widely. Other cases might be cited where techniques for testing the assumptions in an assay have been contributed by the biometrician associated with the project.

Biological inspection tests. Toxicity or safety tests are required for a number of drugs, among them penicillin G. Their purpose is to identify any toxic lots. The present official procedure would be more effective if larger test doses were used. Here again the problem is essentially biometrical. By increasing the dose and "calibrating" each new lot of animals with a standard preparation, any toxicity in a batch of drug could be distinguished from exceptional sensitivity in the test animals (33).

Statistical considerations are also involved in tests of sterility, especially of the antibiotics (33). When the same lots were tested both by a manufacturer and by the Food and Drug Administration, their results on sterility generally agreed in respect to the average percentage contamination for any given manufacturer. However, the lots called "defective" were not the same in the two laboratories. Apparently, the testing procedure was not sufficiently sensitive to separate the defective lots from those which met the sterility standards. Positive identification would require an inordinately large number of tubes. A proposed solution was to test sterility on a grand lot basis, first establishing tolerances, degrees of risk, an acceptable level of contamination greater than zero percent,

and an agreed risk of exceeding that level. This should assure the quality of a manufacturer's product as a whole and provide adequate protection for the public. Here again effective work by biometricians in the pharmaceutical industry could be instrumental in improving official standards.

Bioassays for standardization should be distinguished from adequate inspection or pass-and-fail tests. Biological standardization leads to estimates of potency with a known precision. Pass-or-fail tests would be concerned with checking labelled claims. Although they may use the same laboratory procedures, the latter would have the more stringent requirements before they could meet modern inspection standards. Thus the testing process must be demonstrably in statistical control in respect to the standard deviation and the slope of the dosage-response curve, so that a given difference in the mean response will correspond to a constant difference in potency within a known sampling error. Consumers' and producers' risks must be adopted, together with a specific sampling plan. An operating characteristic curve would be prepared and then the pass-or-fail test would be ready to go. At the present stage of bioassay it seems easier, although possibly less efficient, to assay for potency and pass or fail a lot from its conformity to established tolerances.

Conclusion. From this brief review of the role of a biometrician in the pharmaceutical industry, it should be evident that his services are needed in nearly every stage, in research and development, in clinical testing, in developing biological production processes, and in the control laboratories for biological standardization and control of product quality. The pharmaceutical manufacturer should find him an indispensable member of his team.

References

1. Beecher, H. K. Experimental pharmacology and measurement of subjective responses. Science 116: 157-162, 1952.

2. Beecher, H. K. Pain - controlled and uncontrolled. Science 117: 166-167, 1953.

3. Bliss, C. I. The determination of the dosage-mortality curve from small numbers. Q. J. Pharm. Pharmacol. 11: 192-216, 1938.

4. Bliss, C. I. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26: 585-615, 1939.

5. Bliss, C. I. The first USP collaborative plate assay of penicillin. Mimeographed by the U.S.P., Apr. 1945.

6. Bliss, C. I. Estimation of the error in a clinical assay. Biometrics 8: 237-245, 1952.

7. Bliss, C. I. The Statistics of Bioassay with Special Reference to the Vitamins. Academic Press, N. Y., 1952.

8. Bliss, C. I. and Calhoun, D. W. An Outline of Biometry. Yale Cooperative Corp., 1953.

9. Bliss, C. I., Greiner, Theodore and Gold, Harry. Estimating the dose of cardiac glycoside for human subjects. (In preparation)

10. Bliss, C. I. and Gyorgy, P. The animal vitamin assays. Vitamin Methods, volume II, 41-275, Academic Press, 1951.

11. Bliss, C. I. and Miller, L. C. An analysis of routine penicillin assays from an industrial laboratory. Presented before the Soc. for Stat. Qual. Control, Newark, N. J., March 14, 1946.

12. Bliss, C. I., Pabst, M. L. and Beyer, W. The liver glycogen assay for adrenal cortex extract. (In preparation)

13. Bliss, C. I. and Rose, C. L. The assay of parathyroid extract from the serum calcium of dogs. Am. J. Hyg. Sec. A, 31: 79-98, 1940.

14. Bross, I. Sequential medical plans. Biometrics 8: 188-205, 1952.

15. Chen, K. K., Anderson, R. C. and Robbins, E. B. Gelsemicine, aconitine, and pseudoaconitine: which is the most toxic alkaloid? Q. J. Pharm. Pharmacol. 11: 84-91, 1938.

16. Chen, K. K., Bliss, C. I., and Robbins, E. B. The digitalis-like principles of calotropis compared with other cardiac substances. J. Pharmacol. Exptal. Therap. 74: 223-234, 1942.

17. Davies, O. L. and Hay, W. A. The construction and uses of fractional factorial designs in industrial research. Biometrics 6: 233-249, 1950.

18. Eisenhart, C., Simpson, R. A. and Gillespie, N. A. Ether vs. cyclopropane. Brit. J. Anaesthesia 18: 141-159, 1943.

19. Emmens, C. W. Principles of Biological Assay. Chapman and Hall, 1948.

20. Finney, D. J. Probit Analysis. 2nd Edition. Cambridge Press, 1952.

21. Finney, D. J. Statistical Method in Biological Assay. Griffin; London, 1952.

22. Foster, R. H. K. Standardization of safety margin. J. Pharmacol. Exptal. Therap. 65: 1-17, 1939.

23. Gold, Harry. Personal communication. March, 1953.

24. Greiner, T., Gold, H., Bliss, C. I., Gluck, J., Marsh, R., Matthes, S. B., Modell, W., Otto, H., Kwit, N. T. and Warshaw, L. Bioassay of diuretic agents in patients with

congestive failure. J. Pharmacol. Exptal. Therap. 103: 431-440, 1951.

25. Herwick, R. P., Welch, H., Putnam, L. E. and Gamboa, A. M. Correlation of the purity of penicillin sodium. J. Am. Med. Assoc. 127: 74-76, 1945.

26. Jellinek, E. M. Clinical tests on comparative effectiveness of analgesic drugs. Biometrics 2: 87-91, 1946.

27. Mathes, S., Gold, H., Marsh, R., Greiner, T., Palumbo, F., Messeloff, C. and Pearlmutter, M. Comparison of the tolerance of adults and children to digitoxin. J. Am. Med. Assoc. 150: 191-194, 1952.

28. Mosteller, F. Some statistical problems in measuring the subjective response to drugs. Biometrics 8: 220-231, 1952.

29. Noel, R. A. and Gaunt, W. E. Report on the sterility testing of antibiotic preparations. 1951. Mimeographed.

30. Plackett, R. L. and Hewlett, P. S. Quantal responses to mixtures of poisons. J. Roy. Stat. Soc. B 14: 141-163, 1952.

31. Reilly, J., Chen, G., and Geiling, E. M. K. The action of drug combinations on Plasmodium lophurae in the chick. J. Infect. Diseases 86: 105-109, 1950.

32. Richardson, A. P. Personal communication. May, 1946.

33. Summer Seminar in Statistics. Proceedings of a round table discussion on quality control in pharmaceutical manufacturing. Multilithed by the U. S. Pharmacopoeia, Jan. 4, 1952.

34. Various Authors. The place in statistical methods in biological and chemical experimentation. Consulting Editor E. J. deBeer. Ann. N. Y. Acad. of Sciences 52: 789-942 (Art. 6), 1950.

35. Weil, C. S. Tables for convenient calculation of median effective dose (LD50 or ED50) and instructions in their use. Biometrics 8: 249-263, 1952.

COMPANY EDUCATION ON THE QUALITY PROBLEM

A. V. Feigenbaum
General Electric Company

At base product quality depends upon conscientious people - upon sound, harmonious, effective human effort. This "quality-minded" effort relates to and may be fostered by many factors in a plant. The caliber of employee and labor relations activity, the condition and maintenance of facilities and equipment, the reputation of the plant in its company, community, or industry are all fully pertinent here.

Perhaps most important of all factors which influence quality-minded effort, however, are those of direction, guidance, and example which are brought to bear during the day-by-day, face-to-face relationships between the plant employee and his supervisor. Whether between lathe hand and foreman, foreman and general foreman, superintendent and general manager, the kinds of informal education and guidance provided by the supervisor will, in the final analysis, determine the kinds of quality effort that the man or woman supervised will bring forth. It is in this face-to-face type of relationships that the fundamental type of company education of personnel in quality takes place.

In recent years more formal programs of quality education for plant personnel have been developed as a supplement to the educational relationship that exists between supervisor and employee. These programs have been developed to a degree where major company effort has been expended upon them. They are often characterized by courses of classroom instruction; frequently they include actual training in the operation and utilization of equipment basic to quality production. The programs are often considered as a segment of the over-all plant industrial training activity and are carried on either by the plant Quality Control Engineer or for him by the industrial training director.

Experience seems to show that these programs are most effective when they observe principles such as the following:

1. In development of the programs, the line organization is consulted with respect to the scope and kinds of material used. Line people are encouraged to feel that the course work for the employees supervised is being carried on by the Quality Control and training staff as an assistance to the line rather than as a substitute for it.

2. The formal educational programs, in the long run, include and/or involve all levels of personnel including bench hands as well as general management.

3. The programs provide quality education in three major areas:

a. Knowledge of the facts, practices, and theories incident to modern quality methods.

b. Skills necessary for full utilization of these facts and practices in the production and maintenance of product quality in the plant.

c. Attitudes fundamental to positive acceptance of and achievement in in meeting plant quality goals.

4. The programs are based upon recognition that the solutions to industrial problems - therefore, the solutions to quality problems - are always changing; therefore, education in quality can never be considered as being completed.

5. The programs concentrate upon practical, meaningful quality material as they relate to the particular problems of the plant in question.

The objective of formal quality educational programs may be summarized essentially as follows: "to develop for plant personnel - at all levels - that knowledge of, and those skills in and attitudes toward quality which may contribute to production of company products at minimum cost consistent with full customer satisfaction."

In large plants a long-range program of quality training may include any or all of such activities as the following:

1. A brief and general course for first line supervision in modern methods of planning and controlling quality. Such a course often will run three or four sessions, each of an hour or one and a half hour's duration.

2. A general orientation discussion for middle and upper management levels portraying Modern Quality Control as a management planning and control technique. Such a course often involves two or three sessions, each of an hour or one and a half hour's duration.

3. Orientation training in quality for new plant employees. This work is carried on as a part of regular new employee orientation activity.

4. A course in methods and practice of practical Quality Control techniques for inspectors, laboratory men, selected foremen, and others whose day-to-day work requires new and better quality training. Such a course may run from sixteen to twenty sessions, emphasizing engineering, manufacturing, and inspection phases of Quality Control, with general and brief review of the statistical methods that may be involved.

5. A detailed training course for individuals who may become full-time members of the Quality Control organization or whose work and background make detailed training desirable. Such a course involves not only detailed discussion of practical quality techniques and methods, but it also provides a basic knowledge of the statistical methods that may be useful in a Modern Quality Control program.

While the details of such courses vary in scope and content from plant to plant and company to company, there are certain basic similarities among them which illustrate certain uniform practices in quality education throughout the United States.

In essence it is important to recognize that the education of company personnel in quality is carried on both formally and informally; and that formal training programs are most useful when they support rather than substitute for the informal programs of quality education that exist in the face-to-face, day-by-day relationships between supervisor and employee. When this principle is recognized, and when the formal programs are soundly organized and effectively operating, they may afford companies one of their principal factors in fostering conscientious, quality-minded work on company products.

APPLICATIONS OF STATISTICAL QUALITY CONTROL IN THE PHARMACEUTICAL INDUSTRIES

Thomas J. Moran
Wyeth Laboratories Inc.
Antibiotic Division

Maintaining a standard quality of the antibiotics produced for the medical profession has always been the aim of laboratory and production personnel at Wyeth Laboratories Inc.

During the past decade a science has been developed which provides new techniques for controlling quality. This science is based on the application of statistical methods to the interpretation of quantative measurements. Management has recognized the value of these techniques in the pharmaceutical industry; an industry where measurement is the rule rather than the exception. This recognition has enabled our program to move rapidly toward fulfilling the desires of Management and worker alike, building quality into the finished product and doing this in an economical way.

The development of the Statistical Quality Control Program at Wyeth Laboratories Inc. has followed a sequence calculated to meet the needs of our particular organization. Examples illustrating applications will for the most part follow this sequence of development.

APPLICATIONS OF STATISTICAL METHODS IN RESEARCH AND DEVELOPMENT

James H. Matheny
The Texas Company

The topic of this paper concerns the applications of statistical techniques in Research and Development. These techniques are very valuable tools, by which the research worker may sharpen his statements about physical phenomena and may derive the most information from the least amount of experimental data.

A subdivision of statistical techniques is statistical quality control. This powerful tool is proving valuable in many and varied applications in industry, certainly not excluding research and development.

Other phases of statistics have been found to be especially useful in research work. Among these may be cited the design of experiments - essentially a logical varying of parameters to provide the most information from the smallest number of experimental determinations, analysis of variance- actually the basis of much quality control work, and the least squares evaluation of empirical equations, an example of which is treated in this paper.

Modern mathematical statistics does not necessarily require a large body of data. In quality control applications, we think in terms of a small sample, often using the range as a means of determining an approximation of the standard deviation. For example, a statistical analysis of a characteristic of a fuel derived from three 1953 cars of a given make may be useful. Of course, a fleet of nineteen such cars would probably give a better picture of the situation but, within the probability level we have set, the larger fleet would not necessarily show a different result, merely a more reliable result.

As some of you may know, the statistical methods mentioned above may involve considerable numerical computations. Fortunately, this is not too serious, at least from a long range viewpoint. We no longer use logarithms for most calculations; we use machines. The desk calculator is replacing the slide rule where more than three significant figures are desirable. Larger data processing equipment is becoming available. IBM accounting equipment has been applied to many problems and the large "brains" are developing faster than the techniques for their efficient use.

At our Beacon Laboratories, we have an IBM installation, consisting of a mechanical 602A calculator, tabulator, sorter, and auxiliary equipment, devoted solely to the solution of scientific problems. We have set up more or less standard procedures for many statistical operations, e.g. analysis of variance, least squares equations involving several powers of several independent variables. We are now in

the process of converting our data analysis to IBM's Card Programmed Calculator.

In addition to actual calculations, punched card techniques offer many advantages in the simple handling of data. Errors and inconsistancies are spotted and eliminated, and the mere visual inspection of data is expedited.

Inasmuch as a considerable part of the computational work carried on at Beacon concerns the fitting of equations by least squares, a specific example will be discussed in some detail. This problem is not mathematically difficult, and, as it is synthesized, it is relatively simple computationally.

Specifically, it is desired to evaluate the blending octane number of one component A of a binary blend having this common component. Component A might be benzene, a compound which cannot be satisfactorily run in a pure state in an anti-knock test engine. We will assume that octane number varies linearly with composition. Although this may not be strictly valid, there might be sufficient linearity for blends rich in A to make this calculation realistic. The data are presented in Table I.

We wish to find the common parameter a in the following equations:

$$y = a + b_1 x \qquad n_1 \qquad N = n_1 + n_2 + n_3$$

$$y' = a + b_2 x' \qquad n_2$$

$$y'' = a + b_3 x'' \qquad n_3$$

where the ordinate "y" is octane number and the abscissa "x" is the percent of component B in the binary mixture. From physical considerations, we may assume that the error of measurement in octane number is high relative to the error in composition, so we wish to minimize the sum of the squares of the deviations of the observed octane numbers from the calculated values.

If we treat the binary blends independently using the well-known formulas,

$$b = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}$$

and

$$a = \bar{y} - b\bar{x}$$

TABLE I

OCTANE NUMBERS OF BINARY BLENDS OF A WITH B_1, B_2, AND B_3

AB_1		AB_2		AB_3	
% B_1	O.N.	% B_2	O.N.	% B_3	O.N.
10.5	87.4	5.4	92.1	3.3	93.1
55.2	82.5	17.3	90.7	12.8	94.5
38.0	85.0	21.0	90.4	28.7	95.2
		29.8	89.5	44.0	97.2
				58.9	98.3
				100.0	104.7

Σx	103.7	$\Sigma x'$	73.5	$\Sigma x''$	247.7
Σy	254.9	Σy	362.7	Σy	583.0
Σx^2	4601.29	$\Sigma x'^2$	1657.49	$\Sigma x''^2$	16403.63
Σxy	8701.70	$\Sigma x'y$	6631.95	$\Sigma x''y$	24785.74
Σy^2	21670.01	Σy^2	32891.31	Σy^2	56733.72

x = Percent of component B_1 in a binary blend of component A and component B_1.

x'= Percent of component B_2 in a binary blend of component A and component B_2.

x"= Percent of component B_3 in a binary blend of component A and component B_3.

y = Octane Number of the binary blend.

$\Sigma\Delta^2$ = Sum of the squares of the deviations of the points about the line.

$$s = \sqrt{\frac{\Sigma\Delta^2}{\text{Degrees of Freedom}}}$$

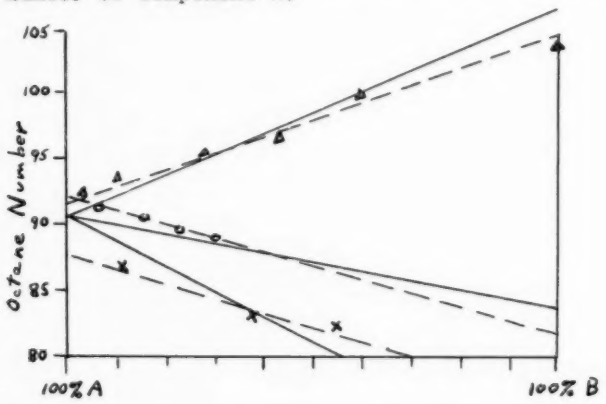
we obtain the following equations:

$$y = 88.68 - 0.1075\,x \qquad s = 0.52$$

$$y' = 92.63 - 0.1064\,x' \qquad s = 0.076$$

$$y'' = 92.37 + 0.1162\,x'' \qquad s = 0.74$$

It will be seen that these lines do not intersect at a common intercept and we do not have a "best" value for the octane number of component A.



The problem of interest is to force the three lines expressing the dependence of octane number on composition to pass through a common intercept on the Y axis. This intercept will then be the best value obtainable from these data for the blending octane number of A.

We form an expression for the sum of the squares of the deviations of the calculated values of "y" from the observed values of "y".

$$\Sigma \Delta^2 = \sum_{i=1}^{n_1} (y_i - a - b_1 x_i)^2 + \sum_{j=1}^{n_2} (y_j - a - b_2 x_j)^2 + \sum_{k=1}^{n_3} (y_k - a - b_3 x_k)^2$$

Setting the partial derivatives of $\Sigma\Delta^2$ with respect to the parameters a, b_1, b_2, b_3 equal to zero, we derive the normal equations:

$$\frac{\partial \Sigma\Delta^2}{\partial a} = 0 \qquad N\,a + (\sum_i x_i)b_1 + (\sum_j x_j)b_2 + (\sum_k x_k)b_3 = \sum_i \sum_j \sum_k y_{ijk}$$

$$\frac{\partial \Sigma\Delta^2}{\partial b_1} = 0 \qquad (\sum_i x_i)a + (\sum_i x_i^2)b_1 = \sum_i x_i y_i$$

$$\frac{\partial \Sigma\Delta^2}{\partial b_2} = 0 \qquad (\sum_j x_j)a + (\sum_j x_j^2)b_2 = \sum_j x_j y_j$$

$$\frac{\partial \Sigma\Delta^2}{\partial b_3} \qquad (\sum_k x_k)a + (\sum_k x_k^2)b_3 = \sum_k x_k y_k$$

We now have the problem of the solution of four linear simultaneous equations in four unknowns, the parameters a, b_1, b_2, and b_3. These have been solved in Table II by a schematic modification of the Gauss elimination method which is used in The Texas Company. We obtain the value of 91.76 for the octane number of A, together with the standard deviation of the points about their respective lines of 1.06.

It is hoped that this brief description of the method of least squares will show the power of the method and also imply the generality of the method. Furthermore, it is shown that computations involving the solution of linear simultaneous equations are practicable with the use of a desk calculator.

REFERENCES

1. Youden, Statistical Methods for Chemists, Wiley, 1951, p. 42.

2. Goulden, Methods of Statistical Analysis, Second Edition, Wiley, 1952.

TABLE II

SOLUTION OF NORMAL EQUATIONS[2] - LEAST SQUARES EVALUATION OF THE OCTANE NUMBER OF A COMMON COMPONENT IN A SERIES OF 3 BINARY BLENDS OF MOTOR FUEL

13	103.7	7.976923075	73.5	5.653846152	247.7	19.05384615	1200.6	92.35384613
	4601.29	3774.083077		-0.1553500106		-0.523540104	8701.70	-0.2319487478
		-586.3038460	1657.49	1150.849999		-1.483608872	6631.95	-0.2537690696
		-1975.883846		-1707.411269	16403.63	8116.407368	24785.74	0.125437316
		-875.3938437		-292.0501335		1018.100359	111295.04	

415.01233632
211.9658284
137.8525378
Σa^2 10.14476135
σ^2 1.1271957
σ 1.06

$y = 91.7566153a - 0.1767896863b_1 - 0.0676691547b_2 + 0.125437316b_3$